The Virtual IC Factory
...can it be achieved?

Technology computer-aided design (TCAD) is the software tool set that allows integrated circuit (IC) technologists to develop new processes and devices without having to build costly test lots. TCAD can potentially provide process engineers the same benefits that electronic circuit CAD (ECAD) has provided to circuit designers. Circuit designers can design, lay out, and test new circuits without fabricating them, and still largely be assured that the circuit will work when manufactured. Will the day come when an engineer can use software to develop a new manufacturing process and predict its performance and yield accurately?

There are three major components to TCAD—small circuit simulation, device simulation, and process simulation. Small circuit simulation is also a part of ECAD, and its best known tool is SPICE. However, a lot of work still needs to be done on the device models contained in SPICE. Device simulation is the computation of the potential, electron, hole, and energy distributions inside a
transistor as a function of the applied bias [1]. This computation is typically done by using finite element approximations to Poisson's and the electron and hole continuity equations. Process simulation is the simulation of each of the manufacturing steps and the computation of the final device structure and doping profiles. It is used to produce transistor geometry and doping profiles, which the device simulators use as input to predict I-V curves and carrier energies, and to generate circuit device models [2].

3-D process simulators remain the weak link in the chain.
three tools are usually linked together to form a TCAD system.

Sub-micron MOS and bipolar devices require three-dimensional simulation. In addition, solving problems with DRAM structures and latchup phenomena necessitate three-dimensional analysis. For example, charge leakage in a DRAM trench capacitor can be controlled by the shape of the oxide in the corner of the trench, an inherently three-dimensional consideration. Three-dimensional device simulators are available from both university and commercial sources. However, three-dimensional process simulation is much less available. Doping profile and material information for device simulation is usually developed from either analytic models or two-dimensional process simulation. Three-dimensional process simulation would be useful not only for analysis of structures, but also in providing accurate doping and material information for use with device modeling.

The accuracy of a device simulation can be no better than the structural input, i.e., “garbage in, garbage out.” Yet, despite the apparent need for three-dimensional process simulation, there are no widely available three-dimensional process simulators, and little activity is focused on their development. Why?

There have been two traditional branches to process modeling. The first is in the simulation of surface evolution and lithography. This activity is topologically two-dimensional, but geometrically three-dimensional. In other words, simulators exist to evolve a two-dimensional surface plane in a three-dimensional space. The other activity is bulk simulation, which focuses on material growth and dopant diffusion. This activity requires the solution of a complex, nonlinear system of equations in the materials of interest. For three-dimensional simulation, it requires three-dimensional grids and structure.

For surface simulation, the state-of-the-art is very good. Excellent results have been attained in modeling the surface evolution. For example, the work of the University of California, Berkeley group on surface evolution during resist development requires advanced computational geometry to evolve the surface without kinks, folds, and loops [3, 4]. An engineering workstation is sufficient for these types of simulations. Figure 1 shows a simulation of corners in an advanced lithography process. In addition to simulation, verification of the simulator is also required. For surfaces, this can also be quite advanced. Figure 2 shows an imager from IBM, which uses a scanning force probe microscope to profile the surface of DRAM trench technology [5]. This technique and others allow new models to be developed and characterized effectively. These characterization techniques can be used to investigate any process that modifies topology—including oxidation, silicidation, lithography, etching and depositing. Measurements of topology are quite advanced.

There are still many challenges in simulation of the wafer surface, but in general these simulators are advanced and commercially available.

The real problem lies with bulk simulation. Bulk process simulation in three dimensions is not widely available. There are four main reasons for this lack of capability. The first major area is measurement techniques. It is not difficult to measure the I-V characteristics of a transistor, but it is costly and less reliable to measure a doping profile. In particular, there are no widely used tech-
niques used to measure doping in two and three dimensions. The second reason is the perceived lack of validity of some of the physical models. The third reason is the cost. Three-dimensional process simulation is thought to be an expensive computation, even more so than device simulation. The final reason is the difficulty in performing process modeling work in universities and company development laboratories.

Challenges

Metrology

One of the most difficult challenges facing one, two, and three-dimensional bulk process simulation is in verifying models. Measurement techniques for investigating doping profiles are not very accurate, and for the most part are inherently one dimensional. This drawback limits the ability to verify the predictions of the models in more than one dimension, and limits the scalability of simpler empirical models. Verification and scalability are further complicated by the fact that all of the available analysis techniques are expensive and difficult to perform.

There are three main techniques available to measure dopant profiles. The first is secondary ion mass spectroscopy (SIMS), which is a destructive technique that makes an etch pit using ion sputtering and examines the backscattered ions. This technique can accurately measure the chemical concentration of the dopant, but can not determine thin dopant layer concentrations very well due to the broadening of the peak. SIMS detection limits place the lower bound on concentration between \(10^{13} \text{ cm}^{-2}\) and \(10^{15} \text{ cm}^{-2}\). measurement of doping levels below this value is usually not accurate. Because the raw SIMS data yields a particle count, the concentration values must be scaled by a dose calibration. This technique can also result in some uncertainty in the profiles. Also, it gives resolution only in the direction of the etch, and therefore is one dimensional. Figure 3 shows a SIMS measurement for a buried boron layer. The SIMS profile shows noise at lower concentrations. Extracting the diffusivity from these curves is difficult, because a large change in diffusivity may result in only a small change in the profile. The error bars on extracted diffusivities are often enormous. This is made worse at low temperatures, because large changes in the diffusivity result in only very small changes in the profile, as mentioned above.

Spreading-resistance profiling (SRP) measures the resistance as a function of depth. SRP measurement systems extract the mobile carrier concentration from the resistance by using mobility models that are modeled as a function of concentration. The mobile carrier concentration can then be related to the doping concentration.

There are several major problems with this technique that limit its accuracy. The first problem is the mobility model, particularly at high concentrations, when damage and precipitates can affect carrier scattering, but may not be accurately included in the mobility model. The second effect is carrier spilling and depletion layers that alter the carrier concentration so that it is different from the doping concentration. These two effects limit SRP to characterizations of moderately doped profiles that are not ultra shallow. Related to SRP is capacitance-voltage profiling, since it also measures carrier concentrations. This technique can be accurately coupled to Poisson solvers to obtain doping profiles, but is limited to lightly doped layers so that they can be depleted by a surface voltage.

The final measurement technique is junction staining. This technique allows chemical delineation of the junction based on the doping type, and is the only simple technique that can be used to examine doping in two dimensions. However, it gives only a single data point—the junction depth. In addition, the degree of staining, and hence the observed junction depth, can vary depending on the chemical solution used. This limits its ability for verification and provides no information on the shape of the profile.

Research on multidimensional measurement techniques is a major concern of the Semiconductor Industry Association’s research roadmap. However, the measurement techniques are still in the research stage and not commonly available for use in characterizing today’s shallow junction technology. This limitation in measurement technology severely hampers development of accurate multi-dimensional simulators. Since it is impossible to verify two-dimensional doping profiles, there is little confidence in the accuracy of the computed profile in two or three dimensions. However, it can be argued that this makes process simulation more important. The only way to examine a doping profile in multiple dimensions is to simulate the manufacturing process.

Models

In device simulation, there are many levels of simulation complexity. There are drift-diffusion simulators, energy balance codes, full hydrodynamic solvers, and finally, Monte Carlo techniques that use a variety of approximations to the band and scattering.
problems [1]. This richness of complexity is not available in process simulation. There have been essentially two approaches to bulk simulation: a phenomenological approach, for example, the decision tree simulator PREDICT [6], and a point-defect-based approach, as demonstrated by SUPREM-IV [7,8]. The phenomenological approach would be natural for three-dimensional bulk simulation since it requires less CPU time, but development of these models has been delayed by the lack of multidimensional measurement techniques of doping profiles.

This leaves full point-defect-based simulators, which offer the promise of accurate multidimensional simulation due to their complete encapsulation of the full physics of dopant diffusion. These simulations rely on accurate calculation of point defect concentrations that are then used to compute the dopant diffusion. Although conceptually it promises great accuracy, there are three main difficulties, model complexity, parameterization, and computation time.

Often there also seems to be sort of an inverse Occam's razor operating in diffusion modeling—Occam's hammer. Occam's hammer states that if a complicating effect can be thought of, it should be included because it will eventually be important as technology shrinks. As the modeling of diffusion has advanced, more and more effects and interactions have of necessity been added to the models in order to adequately represent the forces at work. This increases the burden on parameterization and computation time as additional partial differential equations are solved and parameters are determined.

The parameterization problem can be quite enormous. Figure 4 shows the interstitial diffusivity in silicon as a function of temperature and experimenter. The large scatter in the data does not inspire confidence in the predictive ability of the simulator. Would anyone trust device simulators if we had controversy over four orders of magnitude regarding the carrier mobility? This is essentially the case for point defect based simulations. Both point defect types, interstitials and vacancies, can not be measured directly and can only be measured indirectly through their effect on processes, e.g., diffusion. This leads to circular reasoning—there must be excess interstitials because phosphorus diffuses faster, and the reason for the faster diffusion of phosphorus is that there are excess interstitials present. Part of the scatter in the data is due to different interpretations of the measured result. This reasoning loop is difficult to break, and has resulted in a great deal of controversy about the point defect behavior in silicon.

Finally, there are a large number of equations to be solved in point-defect-based simulators. Because the approach has always been to include as much physics as possible in the simulation to obtain accuracy, it is often necessary to solve 3-5 partial differential equations to obtain the solution for a single dopant profile. This adds to the burden placed on the numerics of the simulation.

**Numerics**

Once the general solutions to the governing partial differential equations are known, they are linearized and then discretized, and placed in a large sparse linear system of equations that must be solved. This linear algebra problem requires a great deal of computation. The computation time increases (as a power law) with the number of discretization nodes, which provides an accuracy versus time tradeoff in the simulation (Fig. 5). As the number of nodes increases, there is an improvement in accuracy at the cost of greatly increased CPU time. Where is the optimal point for a given problem?

There are two related issues to this problem. First, how do we perform the linear algebra so that solutions can be obtained in a reasonable amount of time? Better linear algebra methods allow the CPU time curve in Fig. 5 to be adjusted downward or to have a change in slope. Second, how do we place the nodes to optimize accuracy and minimize solution time? Most codes require the user to select the grid positions, but since the profile changes with time, ideally algorithms would be available to place the nodes in optimal positions. This is complicated by the moving boundaries (oxidation and silicidation) that take place during the solution of the diffusion equations.

Linear algebra is complicated due to the immense problem size involved. A minimum of roughly one hundred grid points is required to represent a dopant profile to 1% spatial accuracy. For three-dimensional simulation, there will be approximately one million nodes. As stated earlier, point-defect-based simulators require three solution variables to solve for a single dopant. The resulting linear sparse system will require approximately three million unknowns. Since solution of linear systems requires approximate solution time on the order of the number of unknowns to the 3/2 power, the computation time required for three-dimensional process simulation is very large.
even for super computers. This situation is similar to that for device modeling, and many of the same iterative techniques can be exploited. Substantial amounts of prior research results exist that can be exploited in solving this problem.

The second problem concerns grid generation and adaptation. Generation is a problem that has been under study for device simulation, and many of these techniques can be adapted for use with process simulation. However, it is important to consider refinement techniques for process simulation, since the profile changes during the simulation. The initial grid may not be adequate for the final structure. The main challenge for grid generation is the adaptation that must occur due to the moving oxide/silicide interface. Oxidation and silicidation are natural companion processes to dopant diffusion simulation, and therefore the grid must be adapted to conform to the growing layer driven by surface reactions. This represents a key problem for 2D simulation reliability, as a large number of the problems with current versions of SUPREM-IV come from the grid adaptation algorithm.

Cultural Barriers
The simplest answer to the lack of three-dimensional process simulation is that there are few research laboratories that are actively pursuing this goal, particularly when compared to the amount of activity in device simulation. Lack of work in the field necessarily produces slower progress. The three major numerical analysis of processes and device conferences (SISDEP, VPAD, and NUPAD) and IEDM all have TCAD papers. Figure 6 shows the distribution by area of the papers presented at the most recent of these conferences. Process simulation and numerical simulation are the smallest areas and are dwarfed by device simulation papers. Many of the numerical papers are device simulation oriented, and roughly half of the process simulation papers are on topology simulation. In the author’s experience, this breakdown is not unusual and reflects the distribution of worldwide effort on process modeling as compared to device, applications, and numerical research. Why is there less research activity in the area of process simulation than device simulation?

The first reason is facilities. Obviously, industrial electronics firms have access to advanced fabrication facilities. However, these facilities are not designed to build process modeling test structures, but instead focus on advanced device structures. The lack of dedicated facilities for experimental work impedes progress. Another barrier is the perceived difficulty in obtaining results. Rather than using resources to characterize and model a particular process, a series of shotgun experiments is run to optimize the structure. In the short term, this approach is probably effective in reducing costs and development time. In the long run, however, little is learned that can be used for truly understanding process optimization and controlling manufacturing for the next generation of technology. Only the largest com...
panies, to date, have been able to afford to invest in process modeling.

On the other hand, most university research labs do not have extensive fabrication facilities and cannot build structures at all. This is in sharp contrast to the situation for device modeling, where transistors are available for measurement from many corporate sources. Fabrication facilities are beyond the means of most universities, and it is increasingly difficult to support those universities that do have facilities. A large fraction of the university based process modeling done in the last decade has come from only a handful of universities which have large, expensive fabrication facilities. This lack of facilities for experimental work translates into a lack of progress.

Finally, the sheer complexity of the problem discourages researchers. In addition to diffusion simulation, the flow equations must be treated for oxide and silicide growth in three-dimensions, and accurate calculation of doping profiles from implantation must be obtained. The wide variety of knowledge and disciplines makes it difficult to form an effective process simulation group.

Possible Solutions

The core of the typical bulk process simulator is organized to facilitate the flow of data between grid, physics, and linear algebra components. User interface and post processing tools are also required. The grid and linear algebra portion of the simulator is directly related to the challenges discussed in numerics, and the physics component relates to metrology and models.

These sections have fairly well defined interfaces, which make it possible to organize and write them independently. Object-oriented languages are a natural development framework for this data encapsulation. The components communicate through elements and stiffness matrices. The physics component needs to be able to assemble the Jacobian for the required partial differential equations for each element in the mesh. In two-dimensional simulation the elements could be either tetrahedra or bricks. The source code for the physics component produces a small, dense Jacobian matrix usually referred to as the stiffness matrix for historical reasons. The dense stiffness matrices can then be copied or summed into the matrix component solver one by one to assemble the large sparse matrix.

If the system is structured in this fashion, it will be possible for a diverse community to actively develop the final software package. Mathematicians can work on sparse matrix techniques and insert their source code with advanced process models and physics. New grid generation schemes can be implemented at the other end of the scale, without affecting the remaining source code components. This has the potential for accelerating development of a three-dimensional process simulator, because each of the major challenge areas can be addressed cooperatively by a wide range of groups in both academic and industrial environments. This in turn tends to lower the structural barriers present to development of a three-dimensional process simulator. For example, several universities (Stanford, California-Berkeley, Michigan, and MIT) have begun collaborating on joint development of TCAD tools, and have already presented some results.

The SIA roadmap identifies process modeling as a top priority area for research over the next decade. This has also been acknowledged by many researchers—it is generally accepted that TCAD systems are limited by drawbacks related to process modeling. All of the problems mentioned here have been identified by the Semiconductor Research Corporation (SRC) and SEMATECH as areas of importance for research in the future. As resources become available, faster development and more collaborative work will take place to address these problems.

Figure 7 shows an early result of SRC funded research to develop three-dimensional bulk process simulation. Arsenic was implanted into the corner of a silicon cube indicated by the mask lines. The surface concentration is plotted. As expected, the three-dimensional corner shows less penetration beyond the mask edge than the two-dimensional sections. The total CPU time for the simulation on a SUN-SPARC 10 was approximately 3 hours, which is quite reasonable. The memory requirements, however, are quite substantial as the simulation required over 200 megabytes of storage. This early result seems quite promising.

Conclusions

Significant strides have been made toward the virtual factory, but the accuracy of the simulation is governed by the weakest link in the chain. The weak link currently is bulk process simulation, in terms of accuracy, valid and comprehensive models, computation time, and available resources and commitment.

![Plot of arsenic concentration at the surface of the wafer. The implant mask is indicated, and the doping contours after annealing are shown.](image)
The accuracy of the simulators is determined by the models and metrology. It is extremely difficult to measure two-dimensional profiles, much less three. Model development, particularly phenomenological models that would be more computationally efficient, are slowed due to the lack of sophisticated measurement techniques. These two factors contribute to lack of motivation for development, since without accurate, verifiable models, three-dimensional process simulators become "garbage in, garbage out."

Numerical problems also provide a major hurdle to development. The first problem is efficient inversion of the large sparse linear system. This problem is similar to the problem faced by device simulation in three-dimensions and can be attacked by some of the same methods. The other problem is grid generation and adaptation. Adaptation is required so that diffusion and oxidation/oxidation problems can be solved at the same time. There has been little work on this topic for three-dimensional simulation.

Finally, there is widespread lack of effort in part due to lack of facilities and apparent benefit for development of three-dimensional process simulation. Overcoming the first three hurdles will help make three-dimensional process simulation more attractive. A possible solution to some of these problems is dividing the program into a set of well-defined interfaces between major modules, so that co-development at several locations is possible. This would allow greater leverage to be applied to the problem, and lower the individual cost of development.

Acknowledgments
I would like to thank the National Science Foundation, SEMATECH, Semiconductor Research Corporation, IBM, and Silvaco for support of my process modeling research. I also thank Kevin S. Jones and Heemyong Park for a critical reading of this article.

References and Further Reading
5. J. Shlinkmann, private communication.

Further Reading


Brain Buster
This problem is being given to eighth graders in Vermont, in these words:

How many different squares of all sizes are on a checkerboard?

Answer to last issue's question:
Let \( g(n) \) be the payoff for \( n \) consecutive tails, and \( p(n) \) be the probability of flipping them. The expected payoff of the game is

\[
E[g(n)] = g(1)p(1) + g(2)p(2) + g(3)p(3) + \ldots = (1)(1/2) + (2)(1/4) + (3)(1/8) + \ldots
\]

\[
= \sum_{n=1}^{\infty} \frac{n}{2^n} = 2.
\]

The fair flipper's fee is therefore two dollars. Now, the expected payoff is

\[
E[g(n)] = (2)(1/2) + (4)(1/4) + (8)(1/8) + \ldots = 1 + 1 + 1 + \ldots
\]

It doesn't make sense intuitively that the fair entrance fee is infinite, but the calculation doesn't lie. The house would never establish this game.

From: The Unofficial IEEE Brainbuster Gamebook, by Donald Mack.

March 1995