Simulation of Defects and Diffusion Phenomena in Silicon

Mark E. Law, George H. Gilmer, and Martin Jaraíz

Introduction
Simulation of front-end processing is a critical component of integrated-circuit (IC) technology development. Today’s electronics are so small that characterization of their material parameters is very difficult and expensive. Simulation is often the only effective tool for exploring the lateral and vertical doping profiles of a modern device at the level of detail required for optimization. Additionally, the cost of fabrication and test lots increases with each technology generation; for this reason, simulation becomes especially cost-effective, if it can be made accurate. Increasingly, process simulation is being performed by harnessing a hierarchy of tools. Ab initio and molecular-dynamics (MD) codes are used to generate insight into the physics of individual particle reactions in the silicon lattice. This information can be fed to kinetic Monte Carlo (MC) codes to establish the dominant, critical mechanisms. Finally, traditional continuum codes can make use of this information and couple with the other process steps to simulate the entire process flow. Both MC and continuum codes can be compared with experiment in order to validate the calculations.

We will introduce the basic concept of the hierarchical modeling of silicon by the illustration of its application to shallow-junction-formation technologies—implantation and diffusion. The article by Cowern and Rafferty in this issue discusses some of the prime experimental evidence; our article will focus on the theoretical framework used to model and understand the data presented in their article.

Ab Initio and MD Methods
Simulations of the processing of silicon are being improved dramatically as more of the operative physical mechanisms are implemented and as more accurate values of the configuration energies and event rates are employed. A large amount of detailed information has recently been provided by atomistic modeling techniques. Quantum calculations of electron distributions (ab initio methods) are the most detailed, and they have an advantage in that the energies of specific atomic configurations can be calculated directly. Furthermore, the electronic properties are obtained as an inherent part of the calculation, providing values for the electrical activity of the defects, for example. They involve a large computational cost and are limited to systems of several hundred atoms. Energy differences between similar configurations are apparently accurate to several tenths of an electronvolt, although some uncontrolled approximations make a definitive calculation of the uncertainties impossible. Figure 1 shows a density contour for electrons in a crystalline Si lattice containing a substitutional B atom. The thicker bonds are connected to the B atom located near the center, indicating a higher electron density in these bonds. During annealing, interstitials and vacancies annihilate one another, and they coalesce to form [311] defects, vacancy voids, and clusters of dopant atoms mixed with point defects. These all have a large impact on the annealing of implantation damage and dopant diffusion. Models of annealing must therefore include detailed information on the energies of clusters of all sizes and compositions. Unfortunately, larger-cluster energies in Si are not easily calculated. Small clusters are more easily simulated, because there are fewer possible configurations.

Accurate properties of a specific configuration can be obtained, but the multiplicity of metastable cluster configurations that must be searched to find the ground state poses a considerable difficulty. Efficient methods are available to relax a given initial configuration to the nearest metastable state. But because of severe limits on the amount of elapsed time that can be simulated, it is seldom possible to “anneal” a cluster of interstitials for sufficient time to reach the ground-state configuration. A search over configuration space is extremely laborious, especially for the larger clusters, simply because of the large number of possible arrangements of the atoms. Recent advances in genetic search algorithms hold promise for the computation of larger clusters, and initial results have been obtained for silicon clusters, in a vacuum, containing up to 20 atoms. The calculations for Si interstitial clusters are more complex, since the crystal lattice induces strain fields, reduced symmetry, and larger barriers between metastable states. For this reason, reliable values for interstitial clusters are probably limited to sizes smaller than about four interstitials. Large clusters that have well-determined [311] structures can be treated, although this entails more empirical methods such as tight-binding, which can simulate a larger computational cell.

The properties of Si interstitials and clusters as calculated by ab initio and other...
techniques indicate that there is a strong, energetic driving force favoring the clustering of individual interstitials into dimers and for further growth into large clusters. Careful \textit{ab initio} calculations give a binding energy of 2.10 eV between the two interstitials in a dimer, for example.\textsuperscript{3} In the case of a large cluster of interstitials in the form of a rodlike [311] defect, classical calculations show that the binding energy of an interstitial to the extended defects is 2.8 eV.\textsuperscript{3,4} This value is smaller than the formation energy of an interstitial, 3.7 eV,\textsuperscript{1} which implies that the [311] defects will eventually anneal out, since they maintain a supersaturated concentration of interstitials in their vicinity.

The simulation of dynamic processes involved in Si implantation and annealing requires a hierarchy of modeling techniques and time scales. The vibrational period of a Si atom in the crystal, about $10^{-15}$ s, sets the “minimum” time period for diffusion events, for example. These events have been simulated using both \textit{ab initio} and classical MD simulations,\textsuperscript{3,5} since small systems are sufficient. Such simulations have a great advantage over attempts to calculate the saddle-point properties, since they represent the actual dynamics of the system of atoms, and all possible diffusion paths are accessible. Classical MD simulations employ empirical force laws, which are adjusted to fit a combination of experimental data and \textit{ab initio} calculations.\textsuperscript{6} These methods are much more efficient than those discussed previously, since the force-law calculation replaces the complex calculations of the electron distribution. Using these methods, it is possible to simulate systems containing tens of thousands of atoms for nanoseconds or more. Accelerated MD methods have been developed to increase the simulation period by several orders of magnitude, making it possible to obtain pathways for infrequent diffusion events.\textsuperscript{7} Often a combination of empirical and \textit{ab initio} techniques can be used effectively; possible low-energy configurations or trajectories could be generated using the simpler method to simulate the dynamics, and the energies could be evaluated by the other technique.\textsuperscript{2,8}

The damage caused by an energetic ion colliding with a Si wafer can also be simulated by MD methods. MD allows simulations over sufficient time and space to contain the energetic collisions of a 5-keV Si ion with a Si target crystal, as illustrated in Figure 2. However, it is not possible to simulate the subsequent anneal used to reduce the damage and activate the dopant, since this process involves times of the order of seconds. A characteristic feature of these simulations is the presence of amorphous pockets, created by the release of large amounts of kinetic energy in local regions or “cascades.” The amorphous pockets appear to account for most of the damage in Figure 2, although several small clusters of interstitials and vacancies are also present. Isolated amorphous pockets recrystallize readily during the ramp-up to the annealing temperature. The amorphous pockets themselves have little direct effect on transient-enhanced diffusion (TED). Simulated annealing of amorphous pockets, using MD techniques, shows that although the amorphous regions recrystallize, small clusters of vacancies and interstitials remain. These are the result of the large displacements that occur during the implant, and these clusters will affect TED.

For purposes of calculating TED, simpler models can be used to calculate ion damage for doses below the amorphization threshold. The MARLOWE simulator calculates all collisions between ions or recoils and the Si target atoms by assuming that the energetic atom interacts only with the closest atom in its neighborhood.\textsuperscript{9} The result of this binary collision is obtained, and based on the amount of energy imparted to the target atom, it will either recoil, leaving a vacant site, or remain in its original site. In this way, the entire collision cascade is calculated. The model provides accurate values for the distributions of implanted dopant atoms, and furthermore yields a distribution of vacancies and interstitials produced by the implantation. Although the amorphous pockets observed in experiments and in MD simulations are not included in the MARLOWE results, a short annealing of either MD or MARLOWE configurations yields small clusters of interstitials and vacancies with similar spatial distributions. For this reason, they both lead to similar amounts of TED.

Monte Carlo Diffusion Simulations

The complete implantation and annealing of a Si wafer takes several minutes, far beyond the time scale accessible to the methods discussed so far. These time scales can be simulated using Monte Carlo (MC) techniques. Here we will describe a hybrid model, using MARLOWE to calculate the damage due to implantation, and an MC diffusion simulator (DADOS) for damage annealing. The coordinates of the vacancies and interstitials are calculated in MARLOWE and transferred to the MC diffusion simulator, where they are selected for diffusion hops at rates based on their diffusivity. Other events included in the diffusion simulation are (1) clustering of like point defects, (2) recombination of vacancies with interstitials, (3) recombination and generation of point defects at surfaces and interfaces, (4) evaporation of point defects from clusters, (5) pairing of point defects with dopant atoms, (6) diffusion of pairs, and (7) clustering of dopant atoms with point defects. We use data from experiments, \textit{ab initio}, together with MC methods to determine the point-defect mobilities, energies, and cluster energies, together with the shapes of the [311] clusters. Because of the limitations of \textit{ab initio} and MC methods, it is absolutely crucial to test the results against actual experiments. The reliability of the energy and rate calculations is not yet sufficient to develop a model based totally on \textit{ab initio} parame-
ters. Some of the parameters developed using information from calculations and experiments are given in Reference 10.

Figure 3 illustrates the concept behind the kinetic MC approach: the figure shows a high-resolution transmission electron microscopy (TEM) view (from Reference 11) of a silicon sample with a [311] extended defect embedded in the silicon atomic rows. In MD, all of the lattice and defect atoms are simulated. In MC, only the atoms belonging to point or extended defects are simulated (represented as circles on the TEM view). In the sample shown in Figure 3, one would see all the lattice atoms vibrating (with a period of about $10^{-13}$ s) and, from time to time (e.g., $10^{-3}$ s), one of the isolated point defects would jump to a neighboring position and be eventually captured by the extended defect. At even longer time intervals (e.g., $10^{-3}$ s), a point defect would be emitted from the extended defect. Since MC only simulates the defect atoms, it starts out with time steps on the order of $10^{-3}$ s, instead of the $10^{-15}$ s step required for \textit{ab initio} or MD calculations. In addition, the fast-moving point defects disappear very quickly, leaving only the extended defects and allowing the time step to automatically be raised to $10^{-3}$ s. Energies obtained from \textit{ab initio} calculations or estimated by fitting experimental data, are used in the MC simulator to determine each event rate and decide which event to perform next. In addition to point defects, the MC simulator includes models for a variety of defect types, including surfaces, clusters, and agglomerates. The main feature of this type of approach is that it can simulate time intervals comparable to real processing times, while using atomic-scale parameters that can be obtained from \textit{ab initio} calculations or experiments.

The simulation of an implant is accomplished by alternating the implantation of an atom using the MC diffusion process with the time of the diffusion prior to the arrival of the next ion, determined by the dose rate of the implant.

We now discuss examples of DADOS simulations, emphasizing results that illustrate atomistic effects. At doses of $1 \times 10^{12}$ cm$^{-2}$ and higher, the results of the model are consistent with the “plus-one” model, which postulates that rapid recombination will eliminate the vacancy/interstitial pairs at such an early stage that only the interstitial kicked out by the implanted ion may contribute significantly to TED. DADOS simulations of implantation and annealing are illustrated on the cover of this issue. This shows the high concentration of interstitials and vacancies at the start of the anneal and the [311] interstitial clusters in the later stages.

At lower doses, DADOS simulations show that interstitial/vacancy pairs can make an important contribution to TED. At a dose of $1 \times 10^{12}$ cm$^{-2}$, the average distance between local cascades for 5-keV Si ions is similar to the depth of the cascades below the surface. In this case, the density of vacancies and interstitials is quite low, and a number of point defects from the vacancy/interstitial pairs escape recombination and diffuse to the surface. These point defects can contribute to dopant diffusion if they encounter a dopant atom. Simulations show that the contribution to TED per implanted ion increases by a factor of 20 for very low doses, and the plus-one model greatly underestimates the expected amount of TED. Experiments measuring TED over a range of doses also show considerably higher diffusion at low doses than that predicted by the plus-one model.

Figures 4a and 4b show a plan-view TEM image (from Reference 15) and the corresponding DADOS simulation of a 40-keV, $5 \times 10^{13}$ cm$^{-2}$ silicon implant after a 5-s rapid thermal anneal at 800°C. In addition to accurate quantitative predictions of the time evolution of the total number of interstitials in clusters, this type of simulation also describes the actual geometry of the defects. This can play an important role in low-energy implants, where the defect size is comparable to the distance to the surface. Due to the MC simulation methodology, it is extremely simple to implement and test new mechanisms. Simulations of B diffusion using the MD model have yielded results that are in excellent agreement with experiments. The calculated fraction of clustered B, which is electrically inactive, agrees with experiments during the entire annealing process. The MC model also provides the full three-dimensional configuration of defects and TED and can represent actual device geometries.

The complexity of the atomic-level interaction mechanisms involved in today’s semiconductor-device processing is a subject of concern for the microelectronics industry: the 1997 Semiconductor Industry Association Roadmap states that “Continuum physics models are no longer sufficient below 100 nm. Tools are needed for the physical and chemical processes at an atomic level.” For example, fluctuations in the channel dopants have been shown to give rise to both a shift in the threshold voltage (as compared with the continuum
roughly with the number of particles, and are traced. Computation time correlates depends on the number of particles that method is statistically sampling the number of particles traced. Since the MC computation time becomes proportional to the path is traced individually, and the com-

gence is that in MC methods, each particle’s interaction between dopants and defects, and defects and clusters. The main differ-

ence is that in MC methods, each particle’s path is traced individually, and the computation time becomes proportional to the number of particles traced. Since the MC method is statistically sampling the number of atoms in the system, the accuracy depends on the number of particles that are traced. Computation time correlates roughly with the number of particles, and numerical error goes as one over the square root of the number of particles.

Continuum methods approach the system from a different angle. In continuum simulation, the physics is formulated as a series of differential equations for each particle type. Typically, these equations are continuity equations—particle gain or loss depends on recombination and diffusion fluxes. Interstitials, for example, get

the following differential equation:

$$\frac{\partial I}{\partial t} = \nabla D \nabla I_c - K_r I_c I_v - C_i^* C_v^*$$

where $C_i$ and $C_v$ are the concentrations of interstitials and vacancies, $D$ is the interstitial diffusivity, $K_r$ is the bulk recombination rate, and $K_{i31}$ and $K_{v31}$ are the forward and reverse reactions of interstitials with {311} defects. This equation can easily be extended with other sources and sinks from interstitial reactions with other species. Each species in the system needs a different differential equation. To model dopant behavior, five differential equations are needed: one for the total dopant concentration, one for each of the two different dopant–defect pairs, and one each for the two different point defects. This level of model is commonly referred to as a five-stream model for diffusion. Additional differential equations are also required to represent clustered defects, for example, the [311] defect population.

Most of the available process simulators make the local-thermal-equilibrium assumption to simplify the five-stream model. Local thermal equilibrium assumes that the dopant–defect pair concentration is the product of the dopant and defect concentrations and the ratio of the forward and reverse reaction rates. The chemical reactions in the five-stream model are balanced. The dopant–interstitial, dopant–vacancy, and dopant equations can all be summed in the five-stream model to produce a single dopant equation. Similar steps can be done for the defect equations. This results in a set of three differential equations for the diffusion system, the so-called three-stream model. The system of equations can account for electric-field effects, nondilute effects, and stress effects on the equilibrium concentrations and diffusion rates. Additional differential equations are still required to represent clustered defects.

The surface boundary condition for point defects plays an important role in determining the response of the system. In the bulk, vacancies and interstitials can only be created together in equal numbers as Frenkel pairs. Only at the surface can they be created or destroyed independently. Defects can be created at the surface by adjusting a surface kink, and annihilated in the reverse process. There can certainly be recombination of the defects at the surface. In addition, surface reactions that are commonly employed in silicon processing can also alter the consumption and injection of defects. This can handle nearly any case of surface condition, since it includes both a recombination rate and generation term. Frequently, in practice, these terms depend on the process happening at the surface, and are proportional to the surface reaction rates.

The five- and three-stream models both produce a nonlinear, stiff set of partial differential equations. They require several different numerical approaches to be combined in a best effort to solve the equations accurately and with minimal computer

![Figure 4. (a) Plan-view TEM micrograph of a 40-keV, 5 x 10^{24}/cm^2 Si implant exhibiting (311) defects after rapid thermal annealing at 800°C for 5 s. (b) Plan-view simulation for the same conditions.](image-url)
time. Spatial discretization is required to reduce the spatial derivatives to algebraic terms. The area of interest is subdivided into boxes containing a single node. Generating the grid (position of the nodes) is critical to the numerical accuracy of the technique. Typically, this is done by hand, which requires experts. Algorithms for automatic placement are becoming more commonplace. Time discretization is also required, but has been automated for some time. The numerical accuracy of the system depends on the grid, while in MC methods it depends on the number of particles simulated. After discretization, the problem is now a large, nonlinear system of equations. Standard methods can be applied for solving this system of equations.

One main advantage of continuum methods is the ease with which they can be coupled to other processes. For example, oxidizing surfaces inject interstitials and are a strong source. An inert silicon dioxide interface does not inject any defects, but it does allow them to recombine. An oxidation solver can be coupled to a continuum diffusion solver, and the resulting system can be treated. The recombination strength of an oxide interface is a matter of some debate, and varies over some three orders of magnitude.20-22 This makes predictive modeling somewhat challenging. The generation rate is usually taken to be proportional to the growth velocity of the interface.

As discussed previously, TED results from the damage created by ion implantation. During the damage anneal, the excess interstitials and vacancies create an enhancement of the diffusivity that can be quite dramatic. This enhancement lasts until the damage is successfully annealed and removed. Below the amorphization threshold, excess interstitials precipitate into {311} rodlike defects. These {311} defects slowly release interstitials and maintain a constant enhancement until the defects have dissolved. The duration of enhanced diffusion is governed by the {311} defect. Figure 5 shows the concentration of interstitials contained in {311} defects along with experimental data of Moller,21 Agarwal,24 Lilak,25 and Eaglesham.26 Implant conditions are noted on the figure, and the simulation is for a $10^{15}$ silic on implantation. The model implemented is a simple {311} model, accounting only for the number of interstitials contained in the defects.27 More complex models can also be implemented that account for the size and distribution of defects.28 Figure 6 shows the diffusion of a lightly boron-doped buried layer in response to the {311} dissolution. The boron diffusion is enhanced from the excess interstitials released by the implant-created defects. The initial Frenkel pair concentrations do not influence this particular example, since they recombine before traversing to the buried layer. The {311} defects control the interstitials in the boron layer and, therefore, the diffusion. Diffusion enhancements are similar to those reported in the experiments of Packan.29

**Figure 5.** Simulated and experimental results of the number of interstitials contained in (311) defects as a function of annealing time at 750°C. Data are for a $10^{14}$ silicon implant annealed at 750°C.

**Figure 6.** Simulated and experimental transient-enhanced diffusion of boron as a function of time during (311) dissolution at 750°C.

**Future Directions**

Today, most silicon technology is developed with the aid of continuum simulation packages, which are commercially available from several sources. Primarily, these codes offer hard-wired models with a range of adjustable parameters. The chip development companies have to spend significant effort in fitting the available models to their own data, and frequently the most recent models are not available. These barriers have always limited the impact of simulation on technology development.

This situation is rapidly changing. *Ab initio* and molecular-dynamics methods are enjoying more widespread use and are beginning to gain great acceptance. The addition of these tools can reduce the number and variety of free parameters and give excellent insight into the appropriate physics. Both Monte Carlo and continuum methods can be built on top of these computations. Monte Carlo simulations offer fast development for new models and greater computation speed for scaled-down device structures. As the devices get smaller, the number of atoms in the structure gets smaller, and Monte Carlo methods can obtain statistical significance faster. Continuum methods are also undergoing rapid change as more modeling power is being delivered to the users. Scripting languages for differential equations are available in several academic tools, and at least one commercial package now offers a modeling interface.

Technology developers will have more powerful simulation tools in the near future to help them design new structures. More reliable physical models will be implemented faster, enabling simulation to make valuable contributions earlier in the design cycle. Simulations will continue to run faster, if for no other reason than increased processor speeds. All of this is happening at a time when complexity of processing is driving the cost of experimental development runs in a fabrication facility upward in an ever-escalating spiral. Computational tools will undoubtedly make a greater impact on future generations of technology.

**References**

As zeolite science continues to grow, many zeolite conferences and meetings are being held around the world, but the premier venue to present new work continues to be the International Zeolite Conferences. The 12th International Zeolite Conference was held in Baltimore, Maryland, in July 1998, and included 774 participants from 37 different countries. It also spawned this comprehensive four-volume, 3360-page proceedings that covers many diverse areas of zeolite study, some of which were only touched upon at previous conferences.

Topics include:

- Diffusion
- Adsorption
- Theory and Modelling
- Structure
- Industrial Applications
- Synthesis
- Catalysis
- Membranes
- Mesoporous Materials
- Solid-State Applications
- Characterization

For complete Table of Contents listing, check out the MRS Web site: www.mrs.org/publications/zeolite/

1999 • Hardbound • 4 Volumes • 3360 Pages

Proceedings of the 12th International Zeolite Conference

$220.00 MRS Members
$245.00 U.S. List
$270.00 Non-U.S. List

Prices do not include shipping and handling. Select one option below.

Shipping/Handling Options

- U.S. Orders (shipped UPS) ____________ $12.00 per 4-volume set
- Non-U.S. Orders (shipped surface) _______ $18.00 per 4-volume set
- Non-U.S. Orders (shipped air freight) ______ $55.00 per 4-volume set

Please allow 8-12 weeks for delivery.