Atomistic Modeling of Complex Silicon Processing Scenarios

Martin Jaraiz, Pedro Castrillo, Ruth Pinacho, Lourdes Pelaz, Juan Barbolla, George H. Gilmer[1] and Conor S. Rafferty[1]
Dept. de Electronica, ETSI Telecomunicacion, University of Valladolid,
47011 Valladolid, Spain.
[1]Lucent Technologies Bell Labs, 600 Mountain Ave., Murray Hill, NJ 07974, U.S.A.

ABSTRACT

The level of sophistication reached by today's Si device fabrication technologies has called for new modeling and simulation schemes, capable of handling the wide variety of interaction mechanisms that govern the complex phenomena that can occur at the atomic level. The kinetic Monte Carlo (KMC) technique seems particularly apt for this task. It takes as input basic materials parameters, derived from ab-initio calculations or from experiments, and is capable of carrying out a detailed simulation up to the dimensions and time scales of current ULSI Si device manufacture. In addition, it can accommodate and efficiently simulate complex interactions between multiple dopant and defect types. We explain the approach and show examples of application in both materials processing and device fabrication. Finally, we present the use of some artificial intelligence techniques (namely, genetic algorithms) that look most promising as methodologies that can easy and efficiently be employed to build the extensive KMC parameter database.

INTRODUCTION

As a result of the efforts to make progress into the deep submicron IC technology, silicon processing is facing an increasing level of complexity. And this situation can only worsen as it is pushed closer to the limit of its possibilities, limit that is expected to be reached around the end of this decade. Front end processing, in particular, is trying to extend the use of conventional and well established doping techniques (ion implantation plus furnace/RTP annealing) into this size regime (quarter-micron to deca-nanometer). However, many new effects show up when those techniques, well suited for larger feature sizes, are applied to sub-quarter micron processing. Since many of these effects are concurrent, the interpretation of experiments becomes ambiguous and the use of predictive process simulation becomes almost imperative. Atomistic Kinetic Monte Carlo (KMC) process modeling seems especially apt to fulfill these needs and is, thus, emerging as a most valuable simulation tool in the forefront of advanced materials processing research.

THE PROBLEM: COMPLEX PROCESSING SCENARIOS

Figure 1 shows a cross-sectional view of the simulation of a typical step in current front end processing: the annealing of a low energy (5keV) As implant to form the source/drain extensions of an N-channel MOS transistor. For that simulation to be truly predictive, physical modeling becomes essential because:

1. There are many different species present: Vacancies (V), self-interstitials (I), arsenic, boron, native carbon and oxygen.
2. There are many different interactions: I-I, I-B, I-C, V-O, C-I-O, …
3. Highly non-equilibrium conditions: Pair reactions, …
4. Extended defects: V voids, I {311}'s, dislocation loops, … with emission/capture rates which are size and shape dependent.
5. Minimum thermal budget, just enough to achieve electrical activation.
6. 2D (short channel) and 3D (narrow channel) effects.
7. Local inhomogeneities: dopant discreteness and clusters.

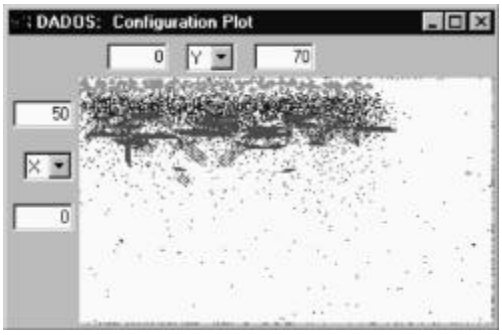

Figure 1. *Cross-sectional view of a simulated S/D extension, where many different physical mechanisms need to be included for the simulation to be predictive.*

SIMULATION APPROACHES

Continuum-type process simulators (which are based on partial differential equations) are still the dominant -if not the only- type of simulator used in the semiconductor industry. However, dealing with the above list becomes practically intractable for such a type of simulator: the number of equations would lead to a prohibitive computation time, leaving aside questions like the size/shape dependent cross-sections of extended defects, to just mention one.

Alternatively, molecular dynamics (MD) can only simulate small regions and -most importantly- very short times, typically less than a nanosecond. This is due to the fact that MD follows the dynamics of every atom -including the vibration of each individual lattice atom- and there is a common timestep for all of them (typically around 10^{-15} s). By introducing a bias potential to accelerate some transitions, it has been possible to extend an MD simulation up to 220 microseconds [1] in a small computation cell (55 moving atoms). However, even that improved MD scheme is still far from being able to simulate a typical processing step.

Instead, the KMC technique seems remarkably suitable to act as a bridge between elemental mechanisms (described at the atomic size and time scales) and the times and sizes involved in typical materials processing steps. In other words, it is possible to, for instance, simulate the front end processing (implants and anneals) followed to fabricate a sub-quarter micron transistor, while still using an atomistic level of description throughout the entire simulation.

THE KINETIC MONTE CARLO APPROACH

Figure 2 illustrates the basic two ideas from which the KMC draws its strength:

1. Only the atoms belonging to defects are considered in the simulation. The "background" lattice atoms are just vibrating around their lattice site position and, thus, need not be included.
2. Instead of a fixed timestep, KMC simulates the sequence of events (point defect emissions, jumps, captures, …) and calculates the (variable) elapsed time between events.

The timestep can, in fact, change automatically from picoseconds to hours and back to nanoseconds, depending on the defects present as the simulation evolves. The sequencing of events and the calculation of the time elapsed between two consecutive events is done based on the current defect configuration (see Ref. [2] for details).

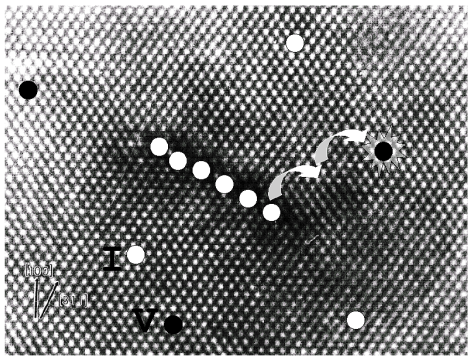

Figure 2. *In the KMC approach, only the atoms in defects (represented as open and closed circles) are followed (TEM courtesy of D. Eaglesham).*

One of the advantages inherent to an atomistic defect description is that the definition of the defect interactions is outstandingly simple and computationally efficient. For instance, the following excerpt from the DADOS source code initiates a V cluster when a V jumps and finds another V within its capture radius neighborhood. If, instead, the interacting particle is an interstitial Boron atom (Bi), then these two particles (jumping V and neighboring Bi) are deleted and replaced by a substitutional Boron:

```
…
   case Vacancy:      // jumping particle is a V
      switch ( neighbor )
      {
            case Vacancy:
            return new Cluster<Vacancy>;
      case Bi:
            this->delete();
             neighbor->delete();
            return new PointD<Boron>;
…
   case Interstitial: // jumping particle is an I
…
```

In this way, different types of interactions between defects can be easily defined. Furthermore, since the 'switch' statement directly jumps to the corresponding 'case' clause, virtually any number of interactions can be included without degrading the simulation speed.

Another distinct feature of the atomistic KMC approach is that, since extended defects are built as agglomerates of individual particles (atoms), the actual 3-D defect geometry can be represented accurately. This, in turn, means that the variation of the capture cross-section can be handled automatically -to a first approximation- as the defect's size and shape evolve. For example, the {311} defects use as capture region the superposition of the capture regions of the constituent self-interstitial atoms. Although this may not be totally accurate, it follows quite closely the growth and shrinkage of the defects and their relative capture strength. In any case, the uncertainties in the binding energies and prefactors will usually dominate over the minor inaccuracies still present in the capture cross-section obtained by the above-mentioned superposition.

SIMULATION EXAMPLES

A variety of examples that exhibit some of the distinctive features of the atomistic

modeling approach can be found elsewhere [2]. Here we will briefly describe two more cases where the KMC approach reveals itself as a unique simulation tool for modeling some special processing effects.

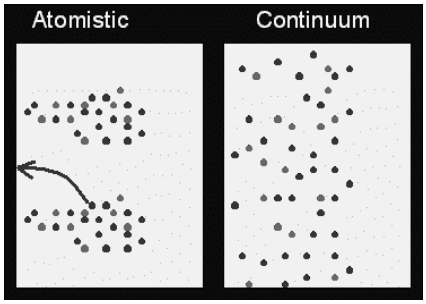

Figure 3. *Two different views of the same low dose implant.*

1. The "+n" Number at Low Implant Doses

As a first example of some of the features that can best -if not only- be described with this type of simulation, Figure 3 illustrates the difference between the atomistic (KMC) and the continuum modeling of a low dose implant (non-overlapping cascades). In both cases the concentration depth profile is the same. However, in the atomistic view, two distinct cascades can be identified, separated by an undamaged region. As a consequence, the balance intracascade/front surface recombination is different from the continuum representation. Figure 4 (from Ref. [3]) shows the dependence of the "+n" number (number of I's left after all V's have recombined, per implanted ion) with the dose. For medium/high doses, the +n is about 1, in agreement with the empirical "+1" rule. However, for low doses the predicted +n number goes up to very high values and then saturates, when the distance between cascades is comparable to the implant range - distance to the surface- because for lower doses the cascades are essentially isolated an only interact with the surface. It has also been observed experimentally that the enhanced diffusivity increases sub-linearly with dose [4].

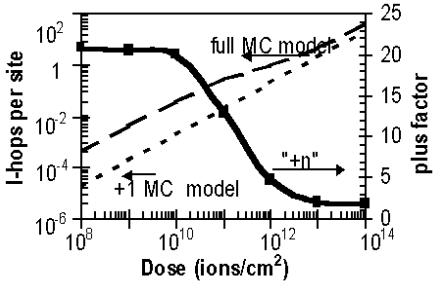

Figure 4. *Dependence of the "+n" number on the implanted dose.*

2. Local Inhomogeneities

The trend to reduce the physical dimensions of devices has come to a point in which the discreteness of the dopants and the granularity introduced by the presence of clusters can play a role in the final operation of the device. The discreteness of the channel dopants, in particular, has been shown to lead to an average shift of the threshold voltage as well as to a source/drain asymmetric behavior in sub-0.1μm MOSFETs [5, 6]. The formation of clusters can, likewise, give rise to local inhomogeneities in the dopant distribution. Figure 5 demonstrates this concept. The left panel (100nm×100nm) shows a typical situation during a high temperature annealing step: {311} defects of different sizes are emitting and capturing self-interstitials but, since the small {311}'s have a higher emission rate than the big ones, there is a local increase around the small {311}'s in the number of self-interstitial hops (right panel) which increases the boron diffusivity and decreases its concentration locally. Atomistic modeling is, again in this case, a unique tool to account for these anomalies in the redistribution of dopants during high temperature processing.

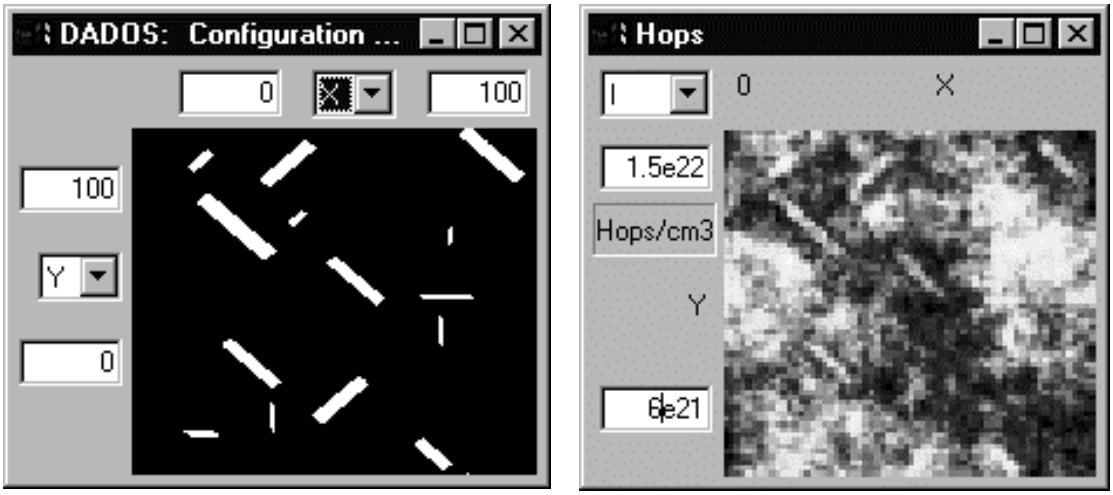

Figure 5. *The small {311} clusters dissolve faster than the big ones (left panel). This leads to a local increase in the number of self-interstitial hops (right panel) that increases the boron diffusivity and decreases its concentration locally.*

OBTAINING KMC PARAMETERS IN COMPLEX SCENARIOS

The KMC scheme can do atomistic simulations with an exceptionally high level of accuracy and detail, and still reach the macroscopic time and length scales involved in some standard processing steps. Furthermore, adding new models is a fairly straightforward task that does not degrade performance noticeably. The main challenge it has to face for modeling complex processing scenarios is that of obtaining the parameters (migration and binding energies and prefactors) for the different species and defects involved. Fortunately, there is a growing awareness in the materials science community towards the exceptional tools that Artificial Intelligence (AI) techniques can nowadays offer to help handle such complex scenarios. Evolutionary Computation (Genetic Algorithms (GA), Genetic Programming (GP), Neural Networks (NN), …) in particular, "is one of the fastest growing areas of computer science" and "it is addressing complex engineering problems that were previously beyond reach" [7]. For example, a GA has been employed to find the minimum energy configuration of carbon clusters up to C_{60} [8]. For that purpose, although there had been many previous attempts to generate the C_{60} buckyball structure from simulated annealing, none had yielded the ground structure [8]. The problem posed to the GA was to find the minimum energy configuration of 60 carbon atoms, starting from random coordinates. Assuming only 10 possible values for each coordinate and ignoring symmetries there are 10^{180} possible configurations to be evaluated (as a reference, the Universe is about 10^{18} s old). Yet, although the GA looks almost like a random search, it found the solution after less than 6000 steps. Minimum energy configurations have also been studied for freestanding Si clusters using this technique [9].

We have used [10] a GA [11, 12] to find the unknown C-related parameters (I_nC_m clusters energies, mainly) that optimize the fit to experimental SIMS profiles. We define (see Figure 6) a chromosome (also called genome or individual) as a particular set of values (alleles) each corresponding to one of the parameters (genes) to be optimized. Each chromosome is tagged with a fitness value that is a measure of how good a solution that chromosome is. The

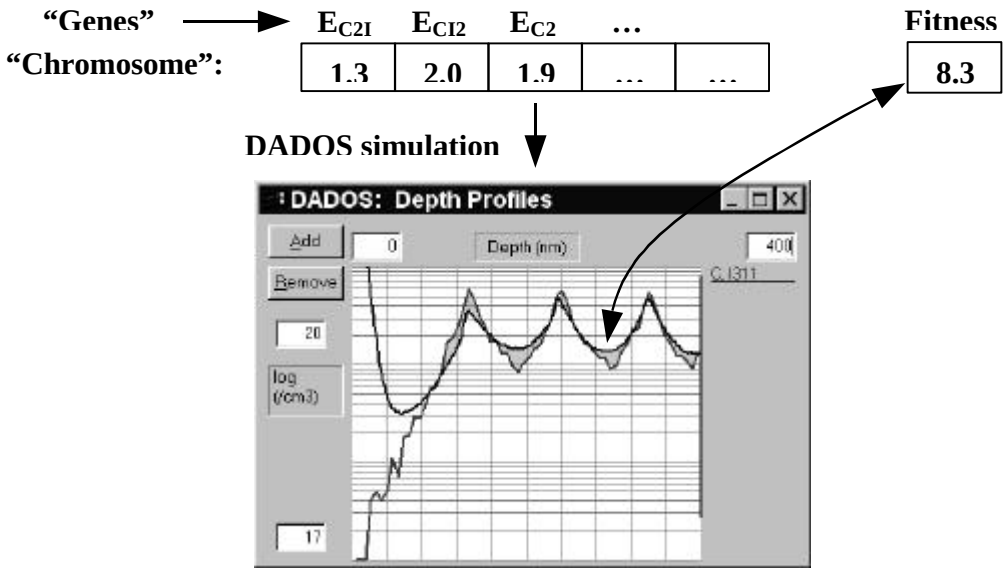

Figure 6. *The fitness of a chromosome (i.e. a particular set of values) is evaluated by running a DADOS simulation and comparing the resulting profile with the experimental SIMS profile.*

chromosome's fitness can be evaluated by running a DADOS simulation and taking the inverse of the area between the resulting profile and the experimental SIMS profile. Those simulations need not be too accurate since we only need a rough (fairly noisy) profile, just enough to be able to decide which chromosome is best when comparing two of them. Instead, we will need to run many (hundreds or thousands) of those simulations and, therefore, they have to be very short. In our case they were 2-3 minutes long, and we needed between 200 and 800 to find the best fit. To achieve such fast simulations, the jump distance was increased (by a factor of 8) and the capture probability was corrected appropriately. The final solutions were always verified by running simulations with the original jump distance (0.345 nm).

Once we have defined the chromosomes and the fitness evaluation we can apply the GA scheme (Figure 7). After the initial random generation of a population (N individuals) two parents are selected for crossover, with a probability proportional to their fitness. Crossover is performed by random exchange of some genes between the two parents. To ensure that the population does not get trapped in a local optimum, there is also a probability of randomly changing some of the genes (mutation). There are, of course, many other possible ways of defining these operations. This is only meant as an example to illustrate the basic concept of GA. The new individuals are then evaluated (fitness), the best N individuals are selected as the new population, and the process is repeated.

Besides its simplicity, the GA approach is totally general in that it works without any specific knowledge of the problem (response surface, gradients, etc.). Other techniques borrowed from AI, like Genetic Programming [13], are also increasingly been used in connection with materials science problems.

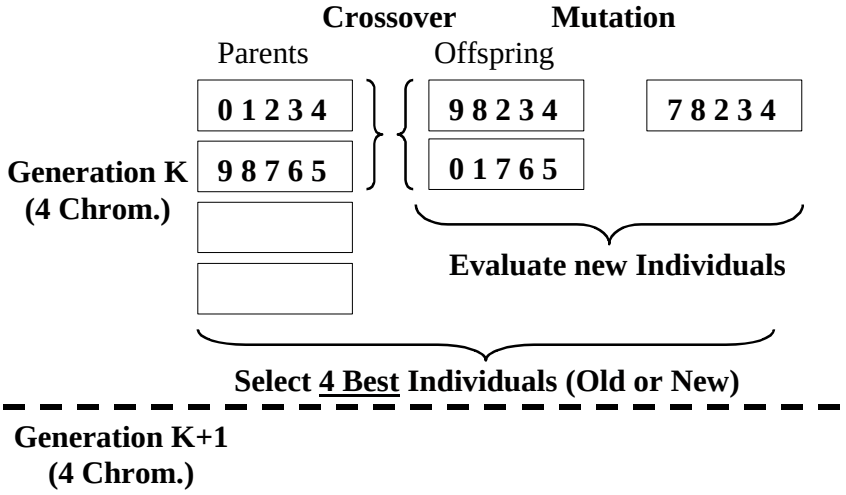

Figure 7. *A Genetic Algorithm scheme.*

CONCLUSIONS

Atomistic process modeling, based on the kinetic Monte Carlo approach, provides detailed and accurate simulations of materials and device processing. In addition, it is fairly straightforward to define new models. It is especially apt for the simulation of complex processing situations where many interacting species and defects are present simultaneously. In such contexts, the limiting step might be the availability of the KMC parameters needed. Fortunately, artificial intelligence techniques (particularly, genetic algorithms), which are increasingly been adopted by the materials science community, can offer invaluable help in the acquisition of those KMC parameters. This synergy, just beginning to be explored, looks most promising for the advancement of materials research.

REFERENCES

1 A. F. Voter, *Phys. Rev. Lett.*, **78**, 3908 (1997).
2 M. Jaraiz, L. Pelaz, E. Rubio, J. Barbolla, G. H. Gilmer, D. J. Eaglesham, H. J. Gossmann and J. M. Poate, *Mat. Res. Soc. Symp. Proc.*, **532**, 43 (1998).
3 L. Pelaz, G. H. Gilmer, V. C. Venezia, H. J. Gossmann, M. Jaraiz and J. Barbolla, *Appl. Phys. Lett.*, **74**, 2017 (1999).
4 P. A. Packan and J. D. Plummer, *Appl. Phys. Lett.*, **56**, 1787 (1990).
5 H.-S. Wong, Y. Taur, *Tech. Dig. Int. Electron Devices Meet.*, 705 (1993).
6 A. Asenov, *Tech. Dig. Int. Electron Devices Meet.*, 223 (1998).
7 D. B. Fogel, *IEEE Spectrum*, p. 26, Feb. 2000.
8 D. M. Deaven and K. M. Ho, *Phys. Rev. Lett.*, **75**, 288 (1995).
9 K. M. Ho, A. A. Shvartsburg, B. Pan, Z. Y. Lu, C. Z. Wang, J. G. Wacker, J. L. Fye, M. F. Jarrold, *Nature*, **392**, 582 (1998).
10 R. Pinacho et al., these Proceedings.
11 K. F. Man, K. S. Tang, S. Kwong, *Genetic Algorithms*, Springer, London 1999.
12 Genetic Algorithm library from *http://lancet.mit.edu/ga/*.
13 W. Banzhaf, P. Nordin, R. E. Keller, F. D. Francone, *Genetic Programming*, Chapter 12, Morgan Kaufmann Publishers, S. Francisco CA, 1998.