Process Simulation

35

Jürgen Lorenz and Peter Pichler

Contents

35.1 Scope and History of Process Simulation . . . 1260

35.2 Simulation of Ion Implantation . . . 1261
35.2.1 Analytical Models . . . 1261
35.2.2 Monte Carlo Simulation . . . 1268

35.3 Simulation of Diffusion and Activation . . . 1271
35.3.1 Intrinsic Point Defects and Impurities . . . 1271
35.3.2 Basic Diffusion and Reaction Mechanisms . . . 1273
35.3.3 Macroscopic Diffusion Behavior of Dopants . . . 1275
35.3.4 Dopant Diffusion at High Concentrations . . . 1278
35.3.5 The Influence of Surface Processes on Dopant Diffusion . . . 1279
35.3.6 Transient Diffusion Effects During Post-Implantation Annealing . . . 1280
35.3.7 Electrical Activation, Clusters, and Solid Solubility . . . 1281
35.3.8 Segregation . . . 1282
35.3.9 Simulation Methodologies . . . 1283

35.4 Simulation of Oxidation . . . 1283
35.4.1 One-Dimensional Simulation of Oxidation . . . 1283
35.4.2 Multidimensional Simulation of Oxidation . . . 1285
35.4.3 Multidimensional Simulation of Stress-Dependent Oxidation . . . 1285

35.5 Lithography Simulation . . . 1286
35.5.1 Basic Principle of Lithography . . . 1286
35.5.2 Principle of Lithography Simulation . . . 1287
35.5.3 Simulation of Imaging . . . 1287
35.5.4 Simulation of Resist Development . . . 1290
35.5.5 State-of-the-Art Lithography Simulation . . . 1290
35.5.6 Examples for State-of-the-Art Lithography Simulation . . . 1290

35.6 Simulation of Deposition and Etching . . . 1291
35.6.1 Outline of Equipment Simulation . . . 1291
35.6.2 Discretization and Movement of Surfaces . . . 1293
35.6.3 Models for Deposition and Etching Rates . . . 1294

35.7 Process Variations . . . 1296
35.7.1 Sources of Systematic Process Variations . . . 1296
35.7.2 Hierarchical Simulation of the Impact of Process Variations . . . 1297

35.8 Conclusions . . . 1297

References . . . 1298

J. Lorenz (✉) · P. Pichler
Modeling and Artificial Intelligence, Fraunhofer Institute for Integrated Systems and Device Technology IISB, Erlangen, Germany
e-mail: juergen.lorenz@iisb.fraunhofer.de; peter.pichler@iisb.fraunhofer.de

Abstract

In this chapter an outline of the very diverse topic of process simulation is given. This includes the process steps of ion implantation and thermal annealing, which introduce, activate, and modify dopant distributions, and the process steps lithography, deposition, and etching, which are used to structure the semiconductor wafer. The section on oxidation deals with the simulation of the oxide growth, whereas the aspect of dopant segregation is included in the section on diffusion. Finally, the impact of process variations is outlined. Overall, the process simulation chapter primarily deals with the physics and the related models for the various process steps, whereas the discussion of generic algorithms, e.g., for the solution of partial differential equations, is left for another dedicated chapter of this book. However, some algorithms which are specific for process simulation are also briefly described in this chapter.

Due to the diversity of the area of process simulation, this chapter could not strive for completeness in the presentation of the physical models. We largely refer to silicon technology, whereas most models can also be applied for or adapted to other top-down semiconductor technologies where, in contrast to bottom-up technologies based on self-assembling, patterning steps, ion implantation and high-temperature process steps are used to generate three-dimensional geometries and dopant distributions.

© Springer Nature Switzerland AG 2023
M. Rudan et al. (eds.), *Springer Handbook of Semiconductor Devices*,
https://doi.org/10.1007/978-3-030-79827-7_35

Keywords

Equipment simulation · Process simulation · Geometry · Dopant distribution · Defect distribution · Point defects · Stress · Process variations

35.1 Scope and History of Process Simulation

This chapter deals with the physical and chemical processes employed to fabricate semiconductor devices based on semiconducting material such as silicon, namely, to generate their geometries and the spatial distributions of dopants, mechanical stress, and potentially further quantities which determine and influence the electrical behavior and the reliability of the devices. The processes to be studied may be grouped into those which primarily influence the geometry, namely, lithography, etching, and deposition, and those which primarily introduce and modify dopant distributions and their activation, namely, ion implantation and thermal annealing. Oxidation primarily changes the local geometry but also affects dopant distributions. Epitaxy is used both to deposit semiconducting layers with desired properties and to introduce dopants, and silicidation is a combination of metal deposition and annealing to form metal-silicon alloys to act as contacts.

Process simulation is an integral part of technology computer-aided design (TCAD). The overall industrial benefit of TCAD in terms of the reduction of development times and costs was already in the 2011 International Technology Roadmap for Semiconductors (ITRS) [1] estimated at about 40%. Although no serious assessment is available on how much process simulation contributes to this benefit, its impact is in any case very significant. Especially, there were also statements from industry that the impact of TCAD cannot be given in percent, because several technological developments would even not be possible at all without the use of TCAD.

Historically and long ago, these processes could be described with macroscopic process parameters, such as the temperature to which wafers were exposed in a furnace reactor, used for annealing or oxidation. During the last decades and in parallel with device scaling, it got increasingly important and difficult to reliably and accurately control such parameters in both space and time. In turn, it has got more and more important to characterize, simulate, and control how such parameters depend on the process equipment used. One of the first effects considered already decades ago was the temperature profiles which occur when a wafer is introduced into a furnace reactor for annealing or oxidation: Whereas the wafer is initially at room temperature, it is gradually heated while being pushed into the furnace reactor. The temperature to which a dopant is then exposed depends on the temperature distribution in the furnace, the movement of the wafer into its target position, and the thermal conduction within the wafer. This is one of the first examples for the need to simulate not only geometries and dopants at transistor scale but also the development of spatial and temporal distributions in the process equipment. This so-called equipment simulation is therefore also considered in this chapter. Especially it is also highly relevant for the advanced fabrication steps used for current nanoscale transistors, such as millisecond annealing, and lithography steps employed to generate patterns far smaller than the wavelength of the light used (e.g., 40 nm structures with 193 nm wavelength double patterning immersion lithography, according to the 2013 International Technology Roadmap for Semiconductors (ITRS) [2]).

With both physical processes and device architectures being driven to their limits, systematic and stochastic variations of the processes used can no more be neglected. The most famous and obvious effect is random dopant fluctuations (RDF) [3]: For example, for threshold voltage implants of 10^{12} cm^{-2}, at mean 4 ions are implanted into a 20-nm-long and 20-nm-wide transistor channel. Because ion implantation is a statistical process, this means that the real number of ions implanted into the channel fluctuates around the mean value of 4. In turn, also the threshold voltage is distributed around its nominal value. Such variations are critical for the overall yield in semiconductor fabrication. A short overview of this aspect is given at the end of this chapter.

The scope of process (including equipment) simulation outlined above is very wide and diverse. In turn, the detailed discussion of the large set of processes and models needed to simulate them would require a dedicated and rather thick book: One of the first reviews of the at that time emerging area of process simulation was published in 1984 as one chapter of a book on the simulation of semiconductor devices in general [4]. Twenty years later, the book on just one (very important) area within process simulation, the modeling of dopant diffusion and activation, published by one of the authors [5] extended to more than 500 pages. Since then, the diversity and complexity of process and equipment models has further grown, due to shrinking features and new processes, materials, and physical effects having come into play. Therefore, this chapter cannot give a comprehensive review of all details. Rather, it intends to give the reader an overview how the input data needed for device simulation, in terms of geometries, active dopants, and strain, can be generated by simulation and which merits and limitations the approaches employed have. In view of this and of the limited space available, we do not include epitaxy, silicidation, and chemical-mechanical polishing CMP. Whereas the chapter

generally refers to silicon as the semiconductor material used, most models can also be applied or adapted to other semiconducting materials. Generally, we refer to specific literature for further details.

Another important aspect to consider is that the level and effort of simulation has historically changed and will continue to do so: Whereas for the simulation of a transistor with one micron channel length, the use of (kinetic) Monte Carlo methods was neither necessary nor possible, a completely different situation applies for sub-decananometer transistors, where shrinking feature sizes both necessitate and partly enable more sophisticated simulation methods. Compared with device simulation, ab initio approaches are hardly used for process simulation, but nevertheless can support the development of computationally more efficient models. During the last decades, simulation has developed from one-dimensional process simulation with academic tools around 1980 [6–8] through two-dimensional process simulation in the 80th and 90th [9, 10] to the current situation where three-dimensional process simulation with commercial tools [11, 12] is the industrial standard. Moreover, in the early days of technology computer-aided design, the individual process steps were mostly treated in isolation, in order to optimize the process step in question, e.g., the annealing steps employed to create high dopant activation while maintaining a shallow junction. In the course of device scaling and more physical models being developed, the interactions between the individual process steps became ever more important. A common example is that dopant diffusion and activation critically depend on the distribution of point defects, which is influenced not only by the annealing step itself but also by other preceding process steps, such as ion implantation or thermal oxidation. In turn, integrated simulation of whole process flows, or the most important part of them, became mandatory. Concerning the transfer of process results into device simulation, in some cases simplistic assumption on device geometry and dopant distributions were made to simplify device studies, e.g., using Gaussian dopant profiles or abrupt junctions. However, such assumptions are sometimes physically not sound and are especially for current nanoscale devices misleading and likely invalidate the results of subsequent device simulations. In turn, the result of the simulation of the whole process flow must be transferred into the device simulator to be used, in most cases necessitating major remeshing. Currently even the gap between the device community and the lithography community is closing, because transistors no more depend on the footprint of the resist structures patterned in lithography steps, but also on the impact of in general non-vertical edges of the resist lines on subsequent etching and deposition processes. This adds to the impact non-vertical nitride, poly, or oxide masks on implanted profiles, which has been included in state-of-the-art process simulation since long.

35.2 Simulation of Ion Implantation

For decades, analytical models have been the workhorse for the simulation of ion implantation. Therefore this chapter starts with this approach and describes it in some detail.

35.2.1 Analytical Models

Some fundamentals of the analytical description of ion implantation can still be taken from the very early book mentioned above [4]. In general, analytical models use distribution functions $f(x)$ which are normalized to unity and vanish for very large (positive and negative) x:

$$f(-\infty) = f(\infty) = 0; \int_{-\infty}^{\infty} f(x)\,\mathrm{d}x = 1 \tag{35.1}$$

The dopant distribution $C(x)$ is then obtained by multiplying the distribution function $f(x)$ with the implanted dose N_D. The part of the vertical distribution for x smaller than zero does not stand for a dopant profile. Rather, it is assumed that the integral across the negative axis corresponds to the amount of particles which are backscattered from the wafer surface during implantation [13].

The analytical models use range moments of the distribution $f(x,y)$, both in vertical (x) and in lateral direction (y), and also partly mixed moments in both directions. Here, x refers to the direction parallel to the incident ion beam and y to the direction perpendicular to this. The first moment in vertical direction is the mean value of the depth in which the ion is stopped, the so-called projected range:

$$R_p = \int_{-\infty}^{\infty} x \cdot f(x)\,\mathrm{d}x \tag{35.2}$$

whereas the further moments in vertical direction, the range straggling ΔR_p, the skewness γ, and the kurtosis β are centered around R_p:

$$\Delta R_p = \sqrt{\int_{-\infty}^{\infty} (x - R_p)^2 f(x)\,\mathrm{d}x} \tag{35.3}$$

$$\gamma = \frac{1}{\Delta R_p^3} \int_{-\infty}^{\infty} (x - R_p)^3 f(x)\,\mathrm{d}x \tag{35.4}$$

$$\beta = \frac{1}{\Delta R_p^4} \int_{-\infty}^{\infty} (x - R_p)^4 f(x)\,\mathrm{d}x \tag{35.5}$$

In case of two-dimensional distributions in x and y, $f(x)$ must be replaced by $f(x,y)$, and the integration runs across x and y.

Due to symmetry, the odd moments in lateral direction are equal to zero. Important lateral moments are the lateral range straggling $\Delta R_{\mathrm{p,l}}$ and the lateral kurtosis β_{l}:

$$\Delta R_{\mathrm{p,l}} = \iint_{-\infty}^{\infty} y^2 f(x,y)\ \mathrm{d}x\,\mathrm{d}y \tag{35.6}$$

$$\imath_l = \frac{1}{\Delta R_{\mathrm{p,l}}^4} \iint_{-\infty}^{\infty} y^4 f(x,y)\ \mathrm{d}x\,\mathrm{d}y \tag{35.7}$$

Because the lateral shape of the implanted doping profile generally depends on the depth x, it is also useful to introduce mixed moments $X^i\,Y^j$ [14]:

$$X^i Y^j = \frac{1}{\Delta R_{\mathrm{p}}{}^i \cdot \Delta R_{\mathrm{pl}}{}^j} \iint_{-\infty}^{\infty} \left(x - R_{\mathrm{p}}\right)^i y^j f(x,y)\,\mathrm{d}x\,\mathrm{d}y \tag{35.8}$$

Whereas again odd moments in y vanish, moments X^iY^2 describe the change of the profile width with depths. Moments X^iY^4 describe the change of the lateral kurtosis with depth.

One-Dimensional Analytical Models

The analytical description of ion implantation is based on distribution functions from statistics, which in the context have no physical background except that ion implantation is a statistical process and that dopant profiles have certain shapes.

Simplest but sometimes still used is the Gaussian distribution, which approximates the vertical distribution as

$$C(x) = \frac{N_{\mathrm{D}}}{\sqrt{2\pi}\,\Delta R_{\mathrm{p}}} \exp\left(-\frac{\left(x - R_{\mathrm{p}}\right)^2}{2\Delta R_{\mathrm{p}}^2}\right) \tag{35.9}$$

It is apparent that due to the non-negligible implantation energy, the depth distribution of implanted ions is not symmetric, and not Gaussian. Generally, Pearson distributions have been shown [15] to best describe vertical profiles of implanted ions. These are defined as solutions of the differential equation:

$$\frac{\mathrm{d}f(x)}{\mathrm{d}x} = \frac{x - a}{b_0 + b_1 \cdot x + b_2 \cdot x^2} \tag{35.10}$$

Here, the parameters a, b_0, b_1, and b_2 depend on the vertical range moments; see Eqs. (35.2), (35.3), (35.4), and

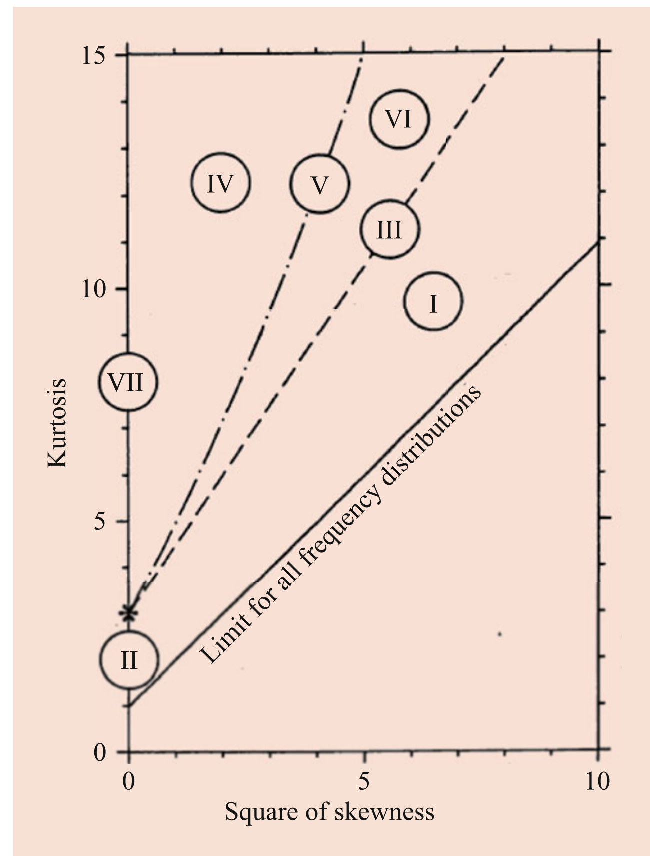


Fig. 35.1 Domains of validity of type (Roman number) of Pearson functions [4] depending on skewness and kurtosis

(35.5) [4]. Depending on the values of b_0, b_1, and b_2, the denominator of Eq. (35.10) may have singularities. In turn, different types of Pearson distributions result, depending on the values of skewness and kurtosis, from which b_0, b_1, and b_2 are calculated; see Fig. 35.1 [4].

For vertical distributions, mostly Pearson IV distributions are suited best:

$$f(x) = K\left[-\left\{b_0 + b_1\left(x - R_{\mathrm{p}}\right) + b_2\left(x - R_{\mathrm{p}}\right)^2\right]^{1/2b_2}. \ \exp\left[-\left(\frac{b_1/b_2 + 2\cdot a}{\sqrt{4\cdot b_2\cdot b_0 - b_1^2}}\right)\cdot \arctan\frac{2b_2\left(x - R_{\mathrm{p}}\right) + b_1}{\sqrt{4\ b_2 b_0 - b_1^2}}\right] \tag{35.11}$$

For vanishing skewness, symmetric Pearson II, Pearson VII, and also Gaussian distributions (for a kurtosis of three) result. Depending on the parameter values, these are used for the lateral dopant distributions.

A very essential aspect in ion implantation is channeling: A larger amount of the ions penetrate deeper into the target if they are implanted along a low indexed crystalline direction.

Generally, implantation is performed with a tilt angle of about 7° to 10° [16], in order to minimize channeling. This situation is described by the Pearson distributions of type IV discussed above. For other (non-standard) tilt angles, range parameters may need to be adapted, depending on the amount of channeling occurring, or Monte Carlo methods should be used, as outlined in Sect. 35.2.2. For implantation into amorphous targets, where no channeling occurs, generally the asymmetric Pearson I and Pearson VI distributions are applicable, with Pearson III and Pearson V as limiting cases. In case of high-dose implantation into crystalline targets, amorphization occurs during the process. This situation can be described by the weighted addition of a Pearson distribution with parameters for a crystalline target and another Person distribution with parameters for an amorphous target [17].

Analytical One-Dimensional Multilayer Models

Ion implantation is not restricted to the ideal cases where the wafer can be divided into areas of bare silicon and areas which are masked by a layer which is not penetrated by the ions. Rather, ions are partly implanted through layers, and mask edges have a finite height and are in general not parallel to the implantation direction. In view of this, models are needed to describe the implantation into multilayer targets. In principle, such models would need to take the differences in the electronic stopping and in the nuclear scattering of the layers into account. Whereas the projected range of the implanted ions is strongly influenced by both electronic stopping and nuclear scattering, the range straggling is primarily influenced by the nuclear scattering as the dominating statistical process in ion implantation. In view of this, some models were published which approximate the change of the doping profile in silicon due to the influence of masking layers by taking the differences of the projected ranges and/or in the projected range straggling in the materials in question into account. The first and well-known model [18] assumed Gaussian profiles for both the mask (layer 1, with thickness t) and the silicon (layer 2) below, using for both layers the projected range and the projected range straggling in question. In addition, the profile in the silicon layer was shifted by

$$d = t \cdot \left(1 - \frac{\Delta R_{p2}}{\Delta R_{p1}}\right) \quad (35.12)$$

The weaknesses of this model were its limitation to Gaussian functions and that the ratio of the projected range stragglings did not sufficiently describe the changes in the stopping powers and, in turn, the profile shift. A generalization of this model [19] removed the restriction to Gaussian profiles:

$$C(x) = C_1(x), x \le t \quad (35.13)$$

$$C(x) = \frac{\Delta R_{p1}}{\Delta R_{p2}} \cdot C_1\left(\frac{\Delta R_{p1}}{\Delta R_{p2}}\, x - t\left(\frac{\Delta R_{p1}}{\Delta R_{p2}} - 1\right)\right), x > t \quad (35.14)$$

However, it still used the modified profile of the first layer also for the second layer and yields wrong results for the limiting case of $t \to 0$, where the profile in the second layer should be $C_2(x)$ and not $\Delta R_{p1}/\Delta R_{p2} \cdot C_1(\Delta R_{p1}/\Delta R_{p2} \cdot x)$.

A model which well considers the original profile shape in the material in question and the changes in projected range due to masking layers is the Numerical Range Scaling model (NRS) [19, 20]. Here, for all layers in the stack, Pearson distributions are used with the range parameters for the material in question. Then, for all but the first layer, the layers on top are considered by shifting the profile based on the thicknesses of the layers on top and the ratios of the projected ranges.

For the implantation through a mask of thickness t (layer 1) into layer 2, this reads

$$C(x) = C_1(x) \text{ for } 0 \le x \le t \quad (35.15)$$

$$C(x) = \alpha \cdot C_2\left(x - t\left(1 - \frac{R_{p2}}{R_{p1}}\right)\right) \text{ for } x > t \quad (35.16)$$

Here, α is a scaling factor after which the model is named: Because no simplifications are made for the profile shapes $C_1(x)$ and $C_2(x)$ (unlike, e.g., in the Ishiwara-Furukawa model [18] where Gaussian distributions are assumed), α cannot be given in analytical form but results from numerical integration, to ensure that $C(x)$ is normalized to the total implanted dose.

A similar scaling of the ranges was also suggested in the LAYER model [21], however restricted to Gaussian profiles, for which dose conservation could be achieved by an analytical scaling factor. In the LAYER model, for the distribution in the second layer, a joint half Gaussian distribution was used, which assumes different range stragglings before and after its maximum, both depending on the projected range in the material of the first layer and projected range stragglings in the materials of the first and the second layer. By comparison with Monte Carlo simulations, it was shown that the range straggling in the second layer does not match especially for very thin or very thick masks [22].

As an extension of the NRS model, in the Improved Numerical Range Scaling model (NRS') [20], the projected range stragglings are corrected based on the layer thickness and the ratios of the projected range stragglings and the projected ranges:

$$\Delta R'_{p2} = \Delta R_{p2} + t\left(\frac{\Delta R_{p1}}{R_{p1}} - \frac{\Delta R_{p2}}{R_{p2}}\right) \tag{35.17}$$

Figure 35.2 shows the comparison between the Numerical Range Scaling model (NRS) and the Improved Numerical Range Scaling model (NRS') and measurements using secondary-ion mass spectroscopy (SIMS) for the implantation of boron with an energy of 100 keV and a dose of 10^{15} cm^{-2} through a silicon nitride layer into silicon. Whereas the NRS model already well describes the shift of the dopant profile in the silicon due to the different stopping power of the nitride layer, NRS' also yields a good description of the profile width in the silicon. However, it still fails to predict the change of the width in case of strange material combinations, such as the implantation of lithium into a stack of cadmium on gold. These are better described by a further refinement of the NRS' model, which was motivated by some aspects of the LAYER model and resulted in a further modified range straggling in the second layer [22]:

$$\Delta R'_{p2} = \Delta R_{p2} + t\left(\frac{\Delta R_{p1}}{R_{p1}} \cdot \frac{R_{p2} + \Delta R_{p2}}{R_{p1} + \Delta R_{p1}} - \frac{\Delta R_{p2}}{R_{p2}}\right) \tag{35.18}$$

The multilayer models outlined above in some way approximate the shift of the doping profile in the buried layers due to the different stopping power of the masking layer or layers and partly also the changes of the profile width. However, during ion implantation backscattering plays an important role. In result, the doping profile in a masking layer may also be influenced by a layer situated below. This holds especially in case of a light masking layer on top of a heavy substrate, and if the maximum of the doping profile is in the second layer, and compared with the projected range not too far from the interface. Based on considerations for the reflection of implanted ions at interfaces and in thin films, a simple model for the reflection at the interface between layers was developed, which is in the following outlined for a masking layer on top of a substrate. Here, the backscattered particles are described by half of a Gaussian distribution, located in the masking layer with its maximum at the interface. The reflected dose equals to a reflection coefficient times the amount of dopants passing the interface, and the projected range straggling of the reflected particles depends on an approximation of the maximum depth of the implantation profile in the masking material, the mask thickness, the implantation energy, and the mean particle energy at the interface. Good agreement with Monte Carlo simulations has been reported even for extreme combinations of implanted ions such as implantation of Li into a stack of silicon on gold [23].

Another effect which is at the edge of the possibilities of an analytical multilayer model is the treatment of channeling. As mentioned above, for implantation into a single semi-infinite layer, the double Pearson approach outlined above yields a good description. In principle, this can be extended by simply applying one of the multilayer models outlined

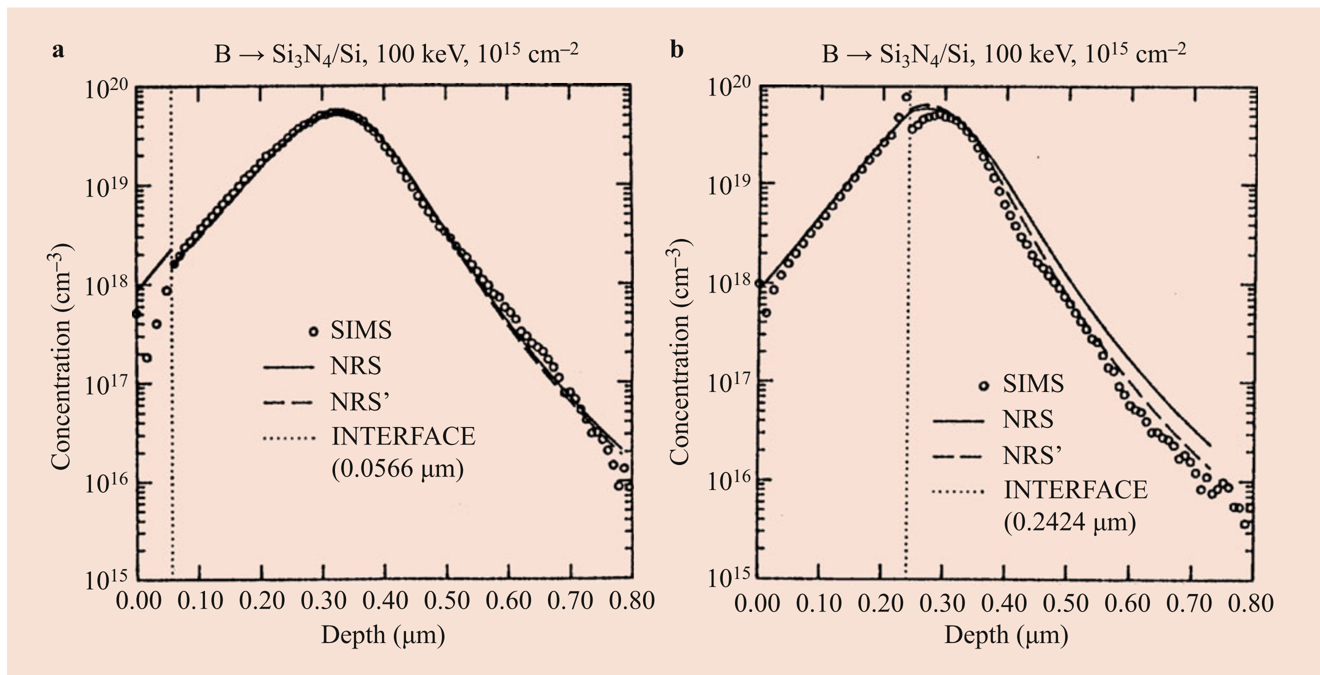


Fig. 35.2 Comparison of the analytical multilayer models NRS and NRS' with SIMS measurements for the implantation of boron at an energy of 100 keV and a dose of 10^{15} cm^{-2} through a Si_3N_4 layer of (**a**) 0.0566 µm; (**b**) 0.2424 µm thickness into silicon [20]

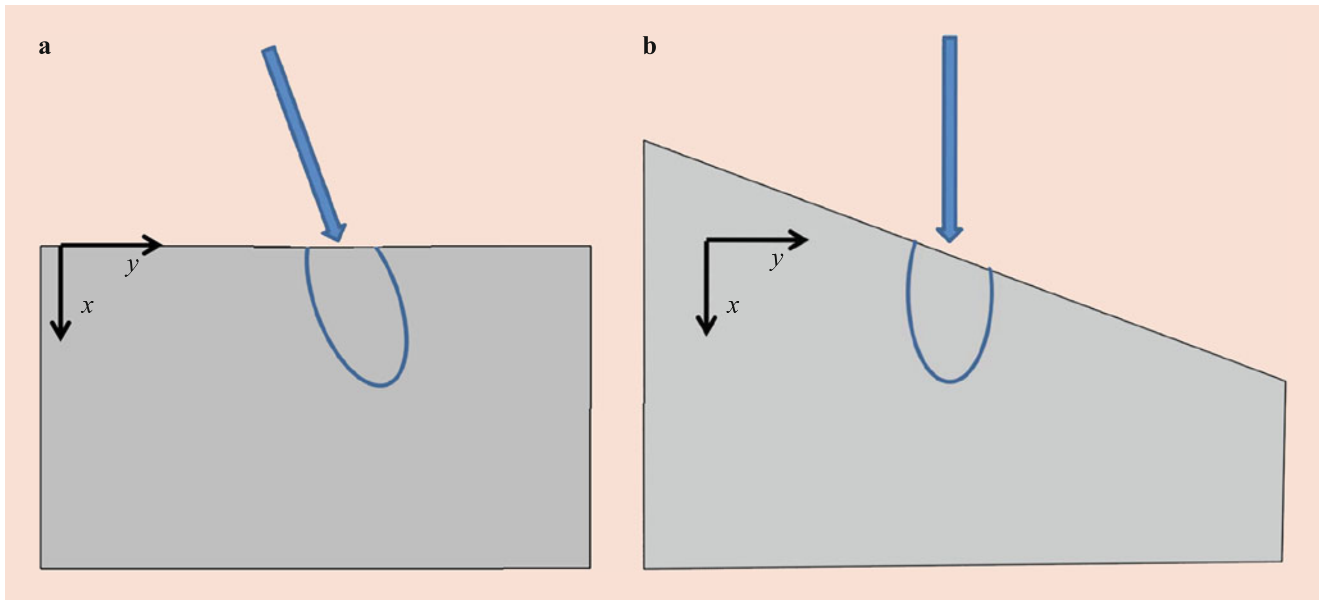


Fig. 35.3 Basic two-dimensional sketch for point response function for titled ion implantation. (**a**) Tilted implantation in standard coordinate system; (**b**) equivalent situation of vertical implantation into tilted wafer

above to both the crystalline and the amorphous Pearson distribution. However, already for the Numerical Range Scaling model explained above, a reasonable agreement with SIMS measurements was reported, and the benefit of applying the double Pearson approach together with NRS was rather limited [22].

In conclusion, in many cases Pearson distributions combined with the Numerical Range Scaling model (NRS) or better the Improved Numerical Range Scaling model (NRS') yield good approximations for one-dimensional dopant profiles. This depends of course on the quality of the range parameters, which must be provided upfront from other sources, as outlined below. The NRS model is also partly being used in commercial process simulation programs [11], however without the modification of the projected range straggling included in the NRS' model. The importance of the use of appropriate model parameters (range moments) is among others highlighted by the availability of different data sets, linking to earlier simulation programs, in current tools. Further refinements of one-dimensional analytical models for ion implantation appeared not to be very useful, due to the growing complexity of description and the need to provide additional parameters which would be difficult to obtain. This would diminish the advantages of analytical implantation models and can be considered as reason why the state-of-the-art in one-dimensional implantation models is about stable since many years.

Multidimensional Implantation Models

The basic approach in all multidimensional implantation models is that first a point response function $g(\boldsymbol{\eta}, \boldsymbol{\zeta})$ is calculated which describes for each point $\boldsymbol{\eta}$ in two or three dimensions the normalized distribution of the ions implanted just at one point $\boldsymbol{\zeta}$ of the wafer surface. Here, it is necessary to take into considerations that the implantation is usually tilted (by about 7°) compared with the wafer normal and rotated out of a two-dimensional simulation plain. This is equivalent to tilting and rotating the wafer geometry by the opposite angles and then considering vertical implantation into this modified structure; see Fig. 35.3. Therefore, in the following "vertical" means in direction of the implantation beam, and "lateral" means perpendicular to the implantation beam.

The two- or three-dimensional dopant distribution then results from the integration across all incident ions beams and multiplication with the implanted dose N_D:

$$C(\boldsymbol{\eta}) = N_D \cdot \int g(\boldsymbol{\eta}, \boldsymbol{\zeta}) \, d\boldsymbol{\zeta} \qquad (35.19)$$

The main topic for multidimensional implantation models is then how to get the point response functions g. The methods range from simplistic analytical assumptions to extracting the point response functions from Monte Carlo simulations; see the corresponding subsection below. Similar to the one-dimensional (vertical) distribution described above, also g must be normalized to unity when integrating across the full two- or three-dimensional space.

The first model published dealt with the implantation through an ideal mask window, where vertical mask edges are situated at $-a$ and a. Strictly speaking the mask should

be infinitely thin (to avoid scattering of ions out of the mask into the mask window), but nevertheless the mask should completely stop ions implanted into the mask. Gaussian shapes of the point response function both in vertical and lateral direction (x and y) were assumed. For implantation at a point $(0,y')$, this reads

$$g(x,y) = \frac{N_D}{\sqrt{2\pi}\,\Delta R_p}\exp\left(-\frac{(x-R_p)^2}{2\Delta R_p^2}\right)\cdot\frac{1}{\sqrt{2\pi}\,\Delta R_{pl}}\exp\left(-\frac{(y-y')^2}{2\Delta R_{pl}^2}\right) \quad (35.20)$$

and the convolution in Eq. (35.19) results in the two-dimensional dopant distribution [24].

$$\begin{aligned}C(x,y) &= \int g(x,y-y')\,dy' \\ &= \frac{N_D}{\sqrt{2\pi}\,\Delta R_p}\exp\left(-\frac{(x-R_p)^2}{2\,\Delta R_p^2}\right)\cdot \\ &\quad \frac{1}{2}\left[\operatorname{erfc}\left(\frac{y-a}{\sqrt{2}\Delta R_{pl}}\right)-\operatorname{erfc}\left(\frac{y+a}{\sqrt{2}\Delta R_{pl}}\right)\right]\end{aligned} \quad (35.21)$$

where erfc is the complementary error function. Approaches like Eq. (35.21) are appropriate in case of device simulation studies which start from analytical dopant profiles. For different mask windows, simply the integration limits need to be changed accordingly. The simplest assumption for implanted and then annealed dopant profiles is to replace the standard deviations ΔR_p and ΔR_{pl} by

$$\sqrt{\Delta R_p^2 + 2Dt} \text{ and } \sqrt{\Delta R_{pl}^2 + 2Dt} \quad (35.22)$$

respectively, because after this replacement, both the point response function of Eq. (35.20) and in turn the dopant profile in Eq. (35.21) satisfy the linear diffusion equation valid for intrinsic diffusion:

$$\frac{\partial C}{\partial t} = D\Delta C \quad (35.23)$$

It should be kept in mind that both Gaussian point response functions for ion implantation and linear diffusion are very crude assumptions. However, assuming Gaussian implantation profiles also in lateral direction, without carrying out the convolution which leads to the complementary error functions erfc in Eq. (35.21), would lead to unphysical results and must be strictly avoided also as input for device simulation.

Two aspects are essential for the generalization of the point response function approach given in Eq. (35.19): First, $g(\boldsymbol{\eta},\boldsymbol{\zeta})$ depends for each entry point $\boldsymbol{\zeta}$ on the local material and on the masking layers on top. Second, it is unphysical to assume that the lateral profile shape, e.g., for the implantation at just one point of the wafer surface, stays constant, independent of the depth: Heavy ions like arsenic are scattered less than light ions such as boron. In result the doping profile somewhat broadens with depth for heavy ions like arsenic, whereas it gets narrower for boron which has a higher likelihood to be backscattered toward the wafer surface. For easier presentation, in the following the two-dimensional case is discussed, whereas the extension to three dimensions is obvious. These effects can be considered by assuming a convolution between a vertical multilayer profile $f(x,y')$ as discussed in the preceding subsection and a lateral profile $g(x',y-y')$ which depends on the depth x':

$$C(x,y) = \int f(x,y')\,g(x',y-y')\,dy' \quad (35.24)$$

Since the vertical distribution $f(x,y')$ is calculated at the lateral coordinate y', the masking effects of the layers through which the ion has passed is taken into consideration. For the lateral distribution $g(x',y-y')$, depending on the range parameters in question, a symmetric Pearson distribution should be used [25]. The vertical coordinate x' can either be equal to x, as in the original publication, or be equal to the entry point into the target for the lateral coordinate y', which is more complicated to implement. The usage of mixed models $X^i\,Y^j$ according to Eq. (35.8) allows to well describe the depth dependence of the lateral projected range [26] and the lateral kurtosis [27]. In the latter paper, a mix of a parabolic and an exponential approach was given for the depth dependence of the square of the lateral range straggling, whereas an exponential approach was presented for the depth dependence of the lateral kurtosis, in both cases depending on the vertical, the lateral, and the mixed range moments introduced in Eqs. (35.2), (35.3), (35.4), (35.5), (35.6), (35.7), and (35.8). Figure 35.4 shows as an example the comparison of the analytical model with Monte Carlo simulations, assuming depth-independent lateral profiles, depth-dependent lateral straggling, and depth-dependent lateral straggling and lateral kurtosis, respectively. In this latter case, the analytical model and the Monte Carlo simulations agree within the statistical error of the Monte Carlo results.

Model Parameters for Analytical Models

As mentioned above, analytical models need suitable vertical, lateral, and partly also mixed range moments. These must be provided upfront as tables depending on the combination of ion and target material, implantation energy, and potentially dose. For the latter two, a suitable number of

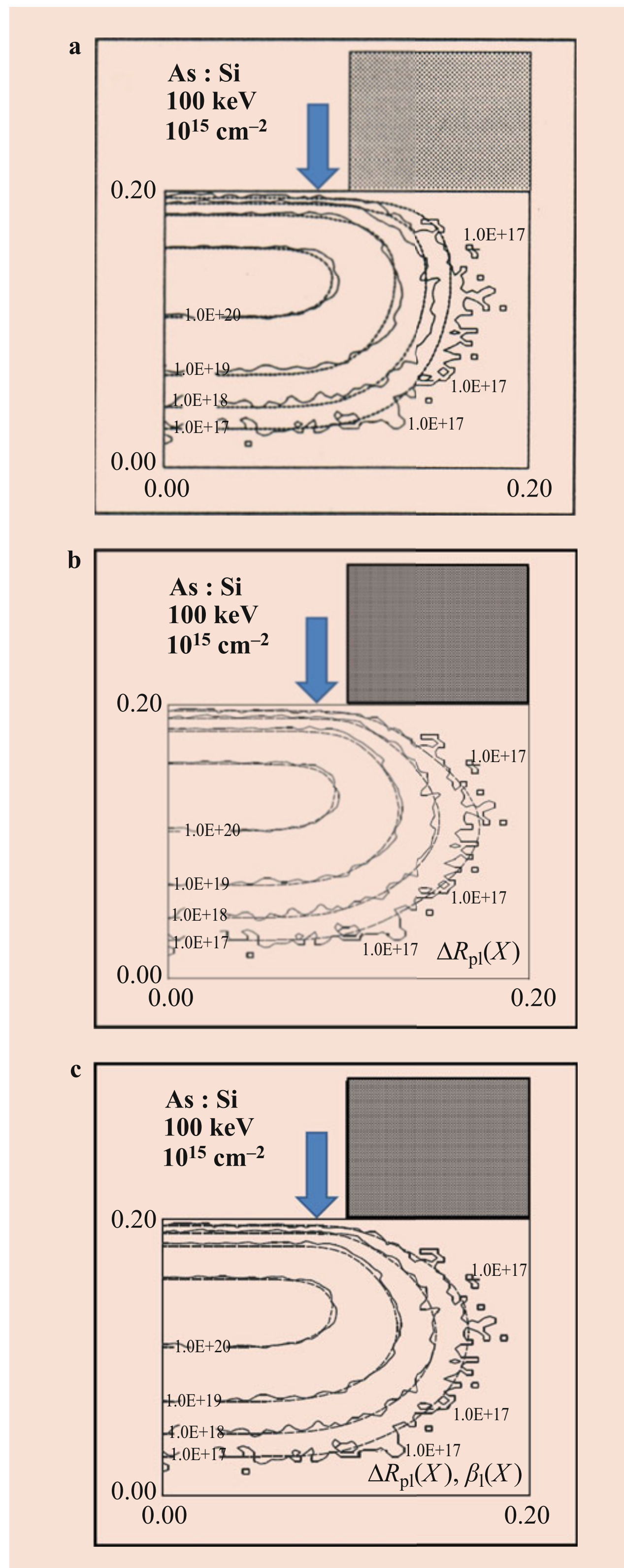


Fig. 35.4 Comparison between analytical model (broken lines) and Monte Carlo simulation of implantation of arsenic at an energy of 100 keV and a dose of 10^{15} cm^{-2} near an ideal mask edge. Equiconcentration lines of 10^{20} cm^{-3}, 10^{19} cm^{-3}, 10^{18} cm^{-3}, and 10^{17} cm^{-3} shown. (**a**) Standard model with ΔR_{pl} and β_l independent of depth x; (**b**) ΔR_{pl} depending on x; (**c**) ΔR_{pl} and β_l depending on x, leading to better agreement for low concentrations

discrete values is used. Whereas state-of-the-art characterization techniques such as SIMS enable the accurate measurement of vertical range profiles and, in turn, the extraction of appropriate vertical range moments, the possibilities for extraction of lateral and especially of mixed range moments from experiment are limited, and the experimental effort becomes excessive. Alternatively, for amorphous targets, all these range moments may be efficiently and accurately extracted from Boltzmann transport calculations, where this equation is not solved directly but its moments are calculated including their dependence on the implantation energy, starting from an infinitesimally small value. This approach was used in the RAMM code [14]. The range moments both for amorphous and for crystalline targets can also be extracted from two-dimensional Monte Carlo simulations, as described below. Here, a sufficient number of particle trajectories must be traced to yield sufficient accuracy of the moments while keeping computation time per implantation energy acceptable for the extraction of tables with various ion/target combinations and implantation energies. Both for the use of Boltzmann transport and Monte Carlo simulations, the use of appropriate physical models for electronic stopping and nuclear scattering, as outlined below, is mandatory. It was demonstrated in the literature [28] that for amorphous targets (vertical), range moments extracted from Boltzmann transport calculations with RAMM and from Monte Carlo simulations using the well-known TRIM code [29] agree well with experimental values obtained by SIMS measurements.

Analytical Models for Intrinsic Point Defects

Because subsequent annealing steps are critically influenced by the distribution of intrinsic point defects (especially vacancies and interstitials of crystal atoms), including those generated during ion implantation, there is a need to also include them in the analytical models for ion implantation. The problem is, however, that the capabilities to measure the distributions of intrinsic point defects are by far not sufficient to extract range moments. In turn, either very simple models are used for point defects or the required parameters have to be extracted from Monte Carlo simulations, as described below. It is also essential to take into account that point defects diffuse very fast and neighboring vacancies and interstitial atoms may partly recombine already during the implantation itself. In turn, in the "+1 model" as the simplest approach, it is assumed that one interstitial and no vacancies are generated per (and at the same position as the) ion implanted.

An established model which is also used in commercial simulators [11] was published already in 1988 [30]. Here, no difference is made between vacancies and interstitials. For the implantation of light ions such as boron and phosphorus, an exponential and a Gaussian function are joined at a position x_0 (naming of variables adapted from [30]), with the

parameters adjusted to assure continuity of the distribution and of its derivative at x_0:

$$(x) = \begin{cases} C_1 \cdot \exp\left(\frac{x}{a_1}\right), x \le x_0 \\ C_2 \cdot \exp\left[-\frac{(x-a_2)^2}{2 \cdot a_3{}^2}\right], x > x_0 \end{cases} \tag{35.25}$$

For heavy ions such as arsenic or antimony, the exponential and the Gaussian distribution are interchanged. The parameters C_1 and C_2 depend on a_1, a_2, a_3, the implantation dose, and the number of point defects generated per ion. In the literature the parameters a_1, a_2, and a_3 were extracted from Monte Carlo simulations, resulting in good agreement between the analytical defect profile and Monte Carlo simulations for the examples presented there. In addition, the lateral defect profile from Monte Carlo simulation could be fitted well with a Gaussian distribution.

For a further brief description of the assumptions used as initial condition for the simulation of post-implantation annealing, see Sect. 35.3.6.

35.2.2 Monte Carlo Simulation

For the Monte Carlo simulation of ion implantation, a large number of ion trajectories is traced until the implanted ions have been stopped, which means that their energy has been reduced to a negligible value. In case of relatively high doses and large structures (e.g., 100 nm scale, as shown in Fig. 35.4), the number of ion trajectories calculated is orders of magnitude smaller that the number of the particles implanted in reality. In consequence, in the simulations pseudo-particles are considered, which have the same properties as the real particles but each represent a large number of them.

The ion trajectories are governed by the energy transfer to the electron gas in the target, which is modeled by the so-called electronic stopping S_e and by the scattering at the (shielded) charges of the crystal lattice, the so-called nuclear scattering S_n:

$$S_{n,e} = -\frac{1}{N}\left(\frac{dE}{dx}\right)_{n,e} \tag{35.26}$$

The first general assumption made is the so-called binary collision approximation, in which only two-body problems are considered, where the impinging ion is at a time only scattered by the potential of one target atom. Because the masses of the implanted ions and of the target atoms are comparable (none of these can be neglected compared with the other), the scattering process is usually calculated not in the laboratory system but in the center-of-mass system. The relationship between implantation energy, masses, and scattering angles in both the laboratory system and the center-of-mass system is given in standard textbooks on mechanics or in standard literature on implantation [31]. First, the reduced mass m_c and the energy E_c of the projectile in the center-of-mass system have to be calculated from its mass m_1, energy E_0, and velocity v_0 in the laboratory system and the mass m_2 of the target atom:

$$m_c = \frac{m_1 \cdot m_2}{m_1 + m_2};\ E_c = \frac{1}{2}\, m_c \cdot v_0{}^2 \tag{35.27}$$

The final results for the scattering angle in the center-of-mass system, Θ, the scattering angle in the laboratory system, ϑ; and energy transfer from the implanted ion to the target ion in the laboratory system are given by

$$\theta = \pi - 2\int_{r_{min}}^{\infty} \frac{p\, dr}{r^2 \cdot \sqrt{1 - \frac{V(r)}{E_c} - \frac{p^2}{r^2}}} \tag{35.28}$$

$$\tan\vartheta = \frac{m_2 \sin\Theta}{m_1 + m_2\cos\Theta} \tag{35.29}$$

$$T = \frac{4\, m_1 m_2}{(m_1 + m_2)^2}\, E_0 \sin^2\left(\frac{\Theta}{2}\right) \tag{35.30}$$

Here, r is the distance between implanted ion and target atom during scattering, r_{min} its minimum value, for which the denominator in Eq. (35.28) equals zero. p is the impact parameter: the minimum distance between the implanted ion and the target in case the ion would not be scattered.

Physical Models

Besides the mechanical problem summarized above, physical models are needed for both the electronic stopping S_e and the nuclear scattering S_n. Because these should be applicable for arbitrary combinations of ions and targets, it is advisable to develop universal models which hold for all combinations. This is generally done by introducing an interatomic screening function $\phi(r)$ [31]:

$$\phi(r) = \frac{V(r)}{Z_1 Z_2 e^2 / r} \tag{35.31}$$

where the numerator is the actual shielded coulomb potential of the target ion and the denominator is the unshielded potential of the implanted ion with charge $Z_1\, e$ and the target atom with charge $Z_2\, e$. As discussed in literature [31], the so-called universal screening potential ϕ_U yields good interatomic potentials, with the Bohr radius a_0 of about 0.529 Å as the length scale:

$$\phi_U = 0.1818\,e^{-3.2x} + 0.5099\,e^{-0.9423x} + 0.2802\,e^{-0.4029x} + 0.02817\,e^{-0.2016x} \quad (35.32)$$

$$x = \frac{r}{a_u} \quad a_u = \frac{0.8854\,a_0}{Z_1^{2/3} + Z_2^{2/3}} \quad (35.33)$$

For the electronic stopping S_e, no universal equation is available. Calculation starts from a fit of the stopping power for protons and includes several scaling and correction steps [31].

Implementation

Current Monte Carlo (MC) simulation programs for ion implantation can be categorized according to the way the target is treated, either as amorphous or crystalline, and whether modifications of the target during implantation are considered or not. Furthermore, several approaches are available to speed up Monte Carlo simulations. Well-known historical examples for MC codes are the amorphous Monte Carlo code TRIM [29] and the crystalline Monte Carlo code MARLOWE [32]. Current MC codes are numerous, including implementations provided by software houses.

Amorphous Monte Carlo simulation: In the simplest case, the target is treated as amorphous, and the effect of modifications during the implantation process on the particle trajectories is neglected. In this case, the true positions of the target atoms are not considered. Employing the binary collision approximation mentioned above, a straight ion trajectory is assumed between the individual scattering effects, which are separated by an assumed mean free path length L. Because most scattering processes occur at impact parameters p which are that large that the scattering angles ϑ (and the transferred energy T) can be neglected, for L a value significantly larger than the interatomic spacing of the (neglected) crystal lattice can be chosen. In parallel, the maximum impact parameter p_{max} to be considered is limited. For consistent simulation, the cylinder with p_{max} as radius and L as length must contain one target atom:

$$\pi \cdot p_{max}^2 \cdot L = \frac{1}{N} \quad (35.34)$$

where N is the density of target atoms. Besides that Eq. (35.34) must be fulfilled, the choice of L and p_{max} depends on the Monte Carlo simulator in question. The actual value of p^2 is then selected at random between 0 (head-on collision with scattering angle 180°) and ${p_{max}}^2$. The scattering angle ϑ and the energy loss T of the ion in the laboratory system are calculated according to Eqs. (35.27), (35.28), (35.29), (35.30), (35.31), (35.32), and (35.33). Between the individual scattering events, the ion energy is reduced by integration of the electronic stopping S_e along the straight path between the scattering events. An approximation for the number of point defects generated is frequently made based on the Kinchin-Pease formula [33, 30], according to which this number is proportional to the ratio of the transferred energy T (or a fraction of this) and the energy needed to generate a Frenkel pair (crystal atom leaving its site and generating a pair of vacancy and interstitial).

Whereas the analytical simulation of ion implantation is in general computationally less expensive, the use of amorphous Monte Carlo simulations is advisable if no suitable range parameters are available, or especially if implanted ions cross interfaces between different layers (including free space) beyond the limitations of the multilayer or multidimensional analytical models summarized above. A common situation where this applies is ion implantation into the neighborhood of a trench. In this case, a significant fraction of the implanted ions is scattered out of the material and into the trench, causing doping of the opposite sidewall of the trench. An example is given in Fig. 35.5.

Crystalline Monte Carlo simulation: The inclusion of the crystal structure increases the program complexity and the computational effort. In essence, based on the current position and direction of flight of the ion, the next neighboring crystal atom is located, or the next few neighboring crystal atoms are identified. In the first case, the binary collision approximation can be used directly, calculating the impact parameter p and subsequently the transferred energy T and the scattering angle ϑ. In the second case in principle, the scattering of an ion in the superposition of two or more screened potentials ϕ_i would need to be calculated, for which in general no analytical formula is available. In turn, this problem is approximated by applying the binary collision approximation separately to each of the scattering processes identified. The details how to extract the final energy transfer T and the scattering angle ϑ for these individual scattering events depend on the simulation code in question. Random parameters are mainly the impact parameter of the first collision of the incident ion with a crystal atom and partly thermal vibrations of the target atoms.

The benefit of the crystalline Monte Carlo simulation, beyond the amorphous Monte Carlo simulation outlined above, is its capability to simulate channeling. For standard implantation into crystalline silicon, where tilt and rotation angles are selected to minimize residual channeling, also analytical models can provide appropriate distributions of implanted ions, provided the models summarized at the beginning of this section are used with suitable range parameters for the combination of implanted ion and crystalline target in question. Such range parameters can be extracted from point response functions obtained from crystalline Monte Carlo

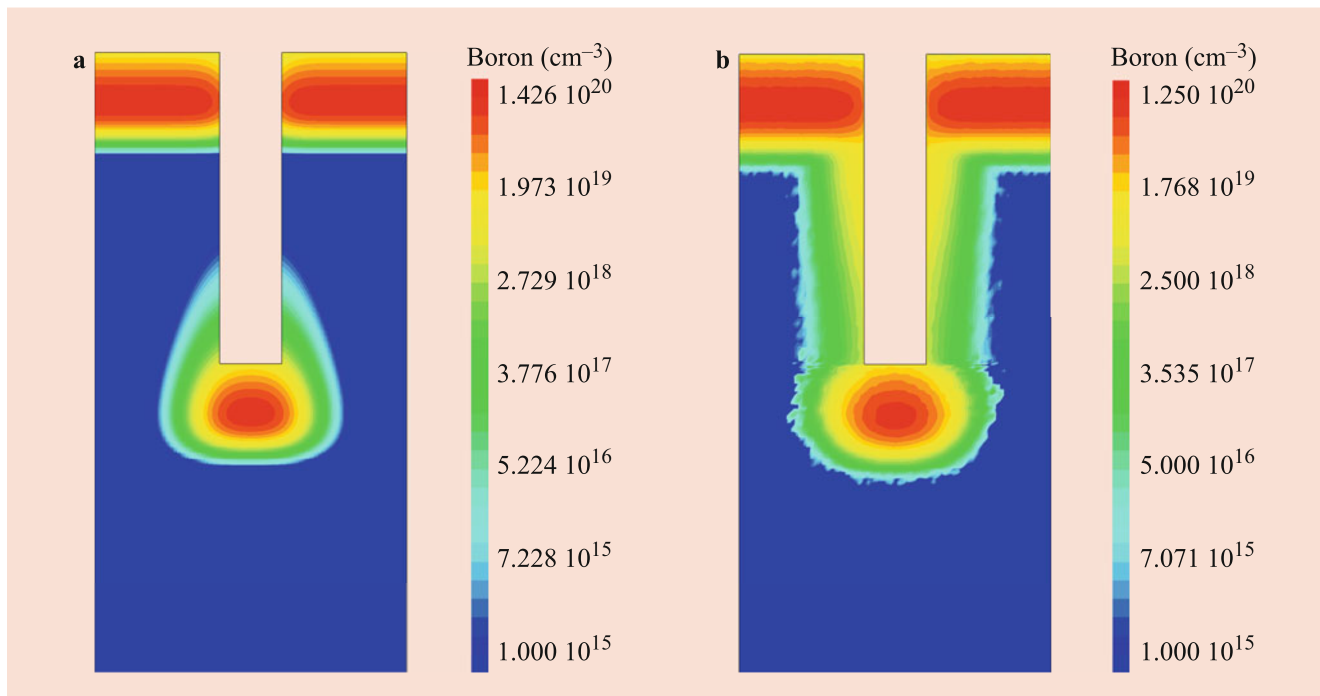


Fig. 35.5 Simulation of implantation of boron at an energy of 20 keV and a dose of 10^{15} cm^{-2} into amorphous silicon with a trench. (**a**) Analytical simulation; (**b**) amorphous Monte Carlo simulation. The doping of the sidewall for the trench cannot be described with analytical methods

simulations or partly from measurements of dopant profiles. On the other hand, advanced materials such as SiC partly show pronounced channeling effects which moreover depend on covering layers of oxide, as illustrated in Fig. 35.6 [34]. The doping spikes occurring here cannot be reasonably described by analytical models.

Dynamic Monte Carlo simulation: The next extension is the inclusion of target modifications which occur when the implanted ion kicks a crystal atom (or an ion implanted before) from its current position. This may even generate cascades of secondary particles. Here, an atom leaves its position in case the energy T transferred to it exceeds a critical value. In consequence, the trajectory of this knock-on particle must be simulated with the same approach as the trajectory of the implanted ion. This increases both the amount of data to be stored and the computational effort. The benefit is that dynamic effects like target modification (e.g., sputtering and amorphization) and the dose dependence of channeling are as well included in the simulations as the extraction of the distributions of the vacancies and interstitials generated.

Statistics: Generally, Monte Carlo simulations require the calculation of a large number of ion trajectories, depending on the dimensionality of the problem and on the size of the simulation domain. For example, for the three-dimensional simulation of source-drain implantation about 100,000 ion trajectories should be calculated to achieve a "smooth" doping profile without artificial fluctuations, resulting in computation times of minutes to hours on a standard PC, depending on ion mass and energy. This situation changes if the simulation domain and the implantation dose are small enough to allow the simulation of the trajectory of each implanted ion: Here, the fluctuations are physically meaningful and part of the results required.

The statistics of Monte Carlo simulations can easily be improved in either of two ways: The first approach is to just calculate a point response function for one surface point and then to copy this on a suitable mesh of surface points, rather similar to the calculation and use of analytical point response functions [35]. Second, in the trajectory split approach [36], the pseudo-particles are split into two or more particles as soon as they enter areas with low concentration of implanted species. The pseudo-particles resulting from these split then represent an accordingly smaller number of real particles and are in turn considered with a lower statistical weight. Their further trajectories are then calculated using random numbers which are independent between these trajectories. This splitting procedure may be repeated several times for decreasing concentrations of the implanted ions. In consequence, additional calculations are only carried out for the relatively rare parts of the trajectories which are situated in low-concentration areas.

Monte Carlo simulation can also be applied to describe novel processes such as plasma doping (PLAD) (also named plasma immersion ion implantation), where charged ions are extracted from a plasma, in order to generate shallow doping profiles with high ion currents. Besides source-drain profiles,

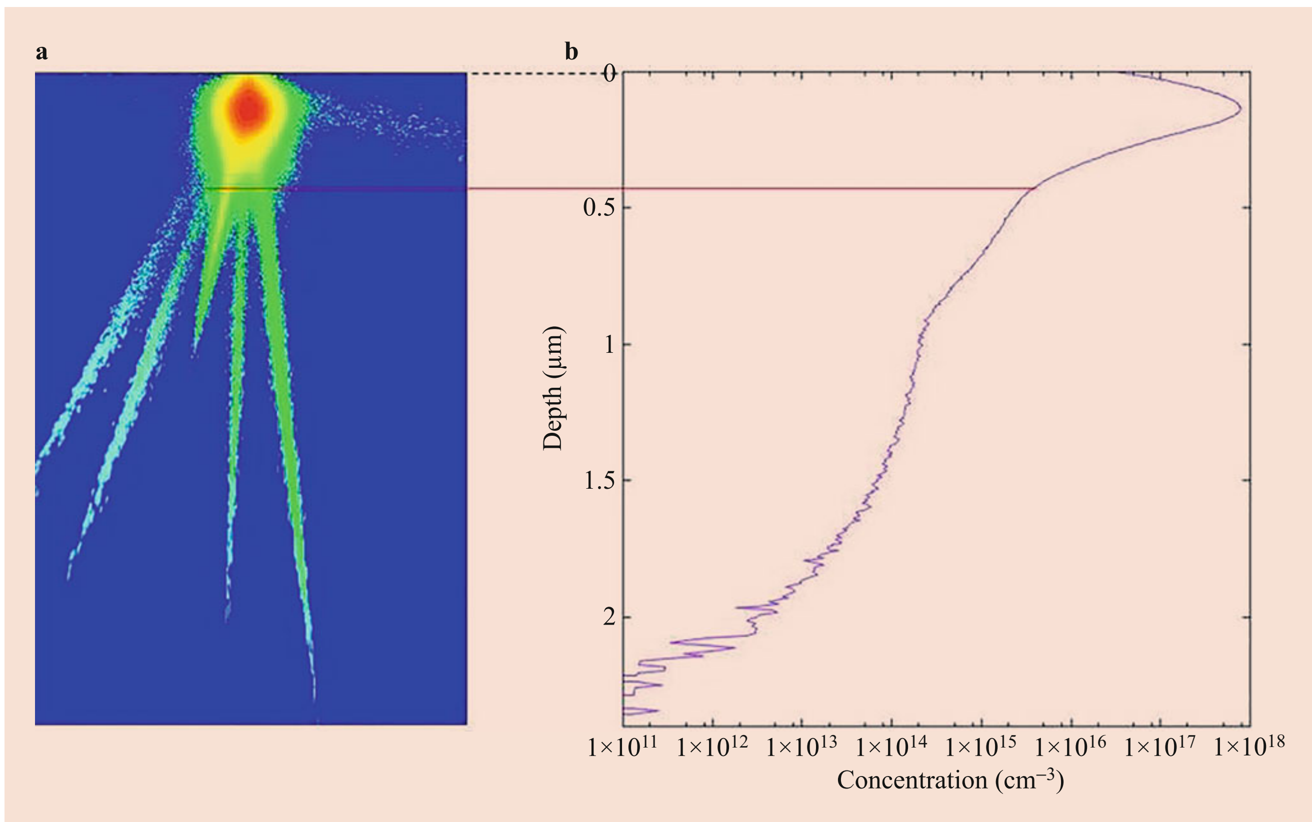


Fig. 35.6 Monte-Carlo simulation of implantation of aluminum at an energy of 140 keV and a dose of 10^{13} cm^{-2} into 4H SiC covered with a 30 nm thick oxide layer: (**a**) 2D point response function and (**b**) 1D vertical profile [34]

PLAD can also be used to generate surface-conformal doping of trenches or FinFETs. For the simulation of PLAD, the energetic distribution of ions can be extracted from physical considerations on the plasma sheath [37] and finally be approximated by analytical expressions, as, e.g., in [38]. This distribution can then be fed into Monte Carlo simulators as described above, which are then used to calculate the spatial distribution of the implanted ions.

35.3 Simulation of Diffusion and Activation

During any process steps at elevated temperatures, the thermally induced movement of dopant atoms and other impurities may lead to a noticeable broadening of their distribution as well as to quasi-chemical reactions with other atoms of the same or other impurity species or with intrinsic point defects. For dopants, the latter typically manifests itself in an electrical activation which is lower than the respective total concentration. Annealing processes at elevated temperatures are used intentionally, e.g., to drive-in dopants deeply into power semiconductors or to electrically activate implanted dopants and to anneal the damage created by the ion implantation processes. However, as indicated already, diffusion and reduced electrical activation may originate from any technological process at high-enough temperatures whether during front-end or back-end processing. In the following, the most important concepts associated with diffusion phenomena in semiconductors are outlined with an emphasis on silicon. For a more comprehensive presentation, the interested reader is referred to reviews such as [5, 39, 40, 41].

35.3.1 Intrinsic Point Defects and Impurities

In crystalline semiconductor materials like silicon, germanium, or SiGe alloys, most of the atoms of which the semiconductor consists are arranged in the form of a diamond structure. In other semiconductors like SiC, the atoms may be arranged in cubic or hexagonal structures. Only few atoms may be missing from their lattice sites or can be found elsewhere. Since such defective atomic arrangements affect only their nearest neighbors, they are referred to as "point defects." Point defects comprising only atoms of which the semiconductor consists of are called "intrinsic point defects." Empty lattice sites, as shown schematically in Fig. 35.7a, are referred to as vacancies. For silicon, even close to the

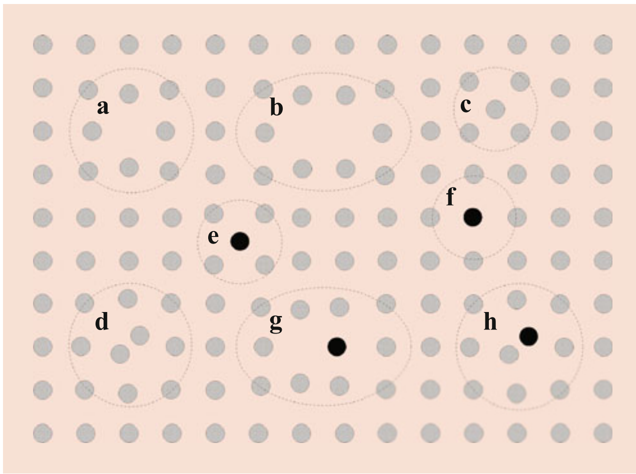


Fig. 35.7 Schematic representation of point defects in elemental semiconductors: (a) Vacancy, (b) divacancy, (c) self-interstitial, (d) interstitialcy, (e) interstitial impurity, (f) substitutional impurity, (g) impurity-vacancy pair, (h) impurity-self-interstitial pair. (after [5])

melting point, less than one out of 10^7 silicon atoms are expected to be missing. Vacancies on neighboring sites (see Fig. 35.7b) are called divacancies. The counterpart of vacancies are semiconductor atoms which either occupy as self-interstitials sites between the lattice atoms as in Fig. 35.7c or share with another semiconductor atom a lattice site as in Fig. 35.7d. Such an atomic configuration is usually referred to as interstitialcy, split-interstitial, or dumb-bell interstitial. Similar to vacancies, less than one out of 10^7 silicon atoms are expected to occupy such interstitial positions even close to the melting point.

Atoms differing from the semiconductor atoms are called impurities. Some of them like oxygen or hydrogen occupy preferentially interstitial sites as indicated in Fig. 35.7e. By far the most important impurities for the Group-IV semiconductors (Si, Ge, SiGe, SiC, etc.) are dopant atoms from the Groups III and V of the periodic system. They preferentially substitute the lattice atoms as shown schematically in Fig. 35.7f. Dopants from Group III require an additional electron to satisfy the four valence bonds to the neighboring silicon atoms which leads to the formation of a hole in the valence band. Based on their ability to "accept" an electron, the Group III atoms are referred to as "acceptors." Substitutional Group V atoms, on the other hand, have one electron too many and can donate it to the conduction band. Accordingly, they are referred to as "donors." In addition to occupying interstitial and substitutional sites, impurities may be found in pairs with vacancies as in Fig. 35.7f or with silicon atoms sharing a regular lattice site as in Fig. 35.7h.

Vacancies can be thought to result from taking one of the lattice atoms from its site and dragging it to a kink at the surface of the semiconductor. Similarly, the intrinsic interstitial defects can be thought to result from dragging semiconductor atoms from a kink at the surface to the respective site. In addition, both can be created as pairs by dragging a lattice atom to an interstitial position and vice versa annihilated by dragging an interstitial atom to a vacancy. The latter processes are usually referred to as bulk generation and recombination of intrinsic point defects.

The probability to find intrinsic point defects in a specific metastable atomic configuration reflects the respective free energy of formation G_X^f. Accordingly, it is not "either-or" but rather "as well as" to find intrinsic point defects in the one or the other configuration. Because of that, it makes hardly sense to discern self-interstitials and interstitialcies, and all such configurations will be referred to as self-interstitials in the following. Besides the point-like configurations, vacancies and self-interstitials were assumed to exist in an extended state, i.e., that one atom in an overall tetrahedrally coordinated region comprising some 10–20 silicon atoms is missing or too much. In the most recent work of Voronkov and Falster [42], as an example, vacancies in three different configurations are postulated to coexist to explain a multitude of seemingly contradictory phenomena.

For the impurity complexes, the situation depends on how the impurities are introduced. For impurities in a closed system, the relative concentrations of the particular atomic configurations need to sum up to the total concentration. For systems in which dopants can enter from a reservoir, their formation energy in the semiconductor will determine the solubility concentration of the impurity in the semiconductor; see Sect. 35.3.8.

Given the concentration of possible sites C_{XS} at which a specific defect can be realized and the number Θ_X of internal degrees of freedom of the defect at a specific site, the concentration of a defect X in steady state is given by

$$C_X^{eq} = \Theta_X \, C_{XS} \exp\left(-\frac{G_X^f}{kT}\right) \qquad (35.35)$$

with k standing for Boltzmann's constant and T for the absolute temperature. For charged defects, G^f will depend on the Fermi level. Deviations of the Fermi level can be expected for dopant concentrations which exceed significantly the charge carrier concentration n_i which would be found under intrinsic conditions, i.e., without dopants. Assuming Boltzmann statistics to be a valid approximation at elevated temperatures, the concentration of a defect in a particular charge state can be expressed in terms of the concentration C_{X^0} of the neutral charge state and the electron concentration n as.

$$\begin{aligned} C_{X^-} &= C_{X^0}\delta_{X^-}\left(\frac{n}{n_i}\right) & C_{X^=} &= C_{X^0}\delta_{X^=}\left(\frac{n}{n_i}\right)^2 \\ C_{X^+} &= C_{X^0}\delta_{X^+}\left(\frac{n_i}{n}\right) & C_{X^{++}} &= C_{X^0}\delta_{X^{++}}\left(\frac{n_i}{n}\right)^2 \end{aligned} \qquad (35.36)$$

Therein, the charge states have been indicated by –, =, 0, +, and ++ for singly negative, double negative, neutral,

positive, and doubly positive. The respective δ_X denote the relative concentration under intrinsic conditions.

35.3.2 Basic Diffusion and Reaction Mechanisms

Simple point defects diffuse by jumping from one energetically favorable site to a neighboring one. In case of simple interstitial impurities as depicted schematically in Fig. 35.7e or h, it is the migrating atom itself which performs the jumps. For the self-interstitial configurations in Fig. 35.7c, d, it could also be always the same silicon atom that moves from site to site. However, it could also be that a self-interstitial assumes the site of a neighboring lattice atom and displaces it to the interstitial site. Finally, the vacancies shown in Fig. 35.7a move when a neighboring semiconductor atom jumps into the vacant lattice site. As long as such point defects do not interact and as long as no forces like electric fields have to be considered, the evolution of their concentration can be described by Fick's laws. Fick's first law

$$J_X = -D_X \operatorname{grad} C_X \tag{35.37}$$

establishes a purely phenomenological definition of the diffusion coefficient D_X of a defect X as a proportionality constant between the net number of defects passing a unit area per time, the diffusion flux J_X, and the gradient of the defect concentration C_X. Fick's second law

$$\frac{\partial C_X}{\partial t} = \operatorname{div}(-J_X) = \operatorname{div}(D_X \operatorname{grad} C_X) \tag{35.38}$$

is a combination of Fick's first law with a continuity equation and constitutes a diffusion equation that can be used to calculate the change in concentration due to diffusion as a function of time t. The concept of a concentration and a gradient obviously requires a certain number of defects per unit volume considered. A complementary view can be obtained by observing the movement of a single defect through the crystal. Given the projection of the mean jump distance L between two stable configurations on one of the axes of the coordinate system, the mean square displacement of the defect from its origin

$$\left\langle D_N{}^2 \right\rangle = NL^2 \tag{35.39}$$

increases linearly with the number N of its jumps. As shown by Einstein more than a century ago, the mean square displacement can be used as an atomistic definition of the diffusion coefficient D in (35.37) and (35.38) in the form

$$D = \frac{\left\langle D_N{}^2 \right\rangle}{2t} = \Gamma \frac{L^2}{2} \tag{35.40}$$

with $\Gamma = N/t$ standing for the number of jumps per time. In general, the diffusion coefficient is a tensor. However, for diamond structures like silicon or germanium, it reduces to a scalar. Transitions between energetically favorable positions require that an energetically less favorable saddle point is assumed and overcome. Diffusive jumps usually require the motion of more atoms than just the moving one. Accordingly, not only the increase in enthalpy H^m at the saddle point has to be considered but also the different entropy S^m of the system. Altogether, the jump frequency is usually expressed as

$$\Gamma = \Gamma_0 \exp\left(\frac{S^m}{k}\right) \exp\left(-\frac{H^m}{kT}\right) \tag{35.41}$$

with the attempt frequency Γ_0. Combining (35.40) and (35.41), diffusion coefficients can be written in the form of an Arrhenius law

$$D = D_0 \exp\left(-\frac{H^m}{kT}\right) \tag{35.42}$$

where the prefactor D_0 comprises all temperature-independent terms. The diffusion coefficient of interstitial oxygen in silicon, as an example, follows this relationship from cryogenic temperatures to the melting point for more than 13 decades [43]. Depending on the species, usual values for the activation energy H^m range from 0 to 5 eV for silicon. To include the effects of mechanical stress on the diffusion coefficient, its temperature dependence (35.42) is completed for hydrostatic pressures P in the form

$$D = D_{P=0} \exp\left(-\frac{P\Delta V}{kT}\right) \tag{35.43}$$

introducing the activation volume ΔV. Positive numbers for the activation volume mean retarded diffusion for compressive stress and enhanced diffusion for tensile stress and vice versa for negative activation volumes. In magnitude, activation volumes are expected to range up to the volume associated with a lattice atom. For non-hydrostatic pressures, a tensor form has to be adopted in which the hydrostatic pressure is replaced by the strain tensor and the activation volume by a strain-activation tensor.

As indicated already above, defects are expected to coexist in a variety of charge states. In case of dopant concentrations exceeding the intrinsic concentration of charge carriers n_i, an electric field will be established in the semiconductor which exerts a force on the charged defects. Accordingly, the diffusive jumps will be biased since the defects gain or lose energy when they move in the field from their energetically favored site to the saddle point. To include such effects, Fick's first law is adapted in the form

$$J_X = -D_X \operatorname{grad} C_X - z_X \mu_X C_X E \tag{35.44}$$

where the charge state z has been defined here as the number of negative charges associated with the charge state (i.e., $+1$ for a singly negatively charged defect or -2 for a doubly positively charged defect). It should be noted, however, that an association of the charge state with the number of positive charges is just as common in the literature. The mobility μ was already shown by Einstein to be related to the diffusion coefficient by

$$D/\mu = kT/q \tag{35.45}$$

with q standing for the elementary charge. Expressing the electric field as the negative gradient of the electrostatic potential ψ, (35.44) takes the familiar form

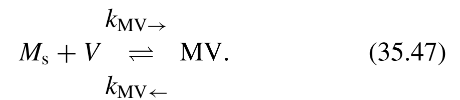
$$J_X = -D_X \operatorname{grad} C_X + z_X D_X C_X \operatorname{grad}\left(\frac{\psi}{U_T}\right) \tag{35.46}$$

with the thermal voltage U_T introduced for kT/q.

With defects co-existing in a number of charge states, any charge state contributes to the diffusion process with a different diffusion coefficient. Adding the contributions of the individual charge states to total concentrations and fluxes, one arrives again at the diffusion Eqs. (35.38) and (35.46). The diffusion coefficients and charge states in these equations are Fermi-level-dependent mean values weighted with the relative concentration of the respective charge state. However, particularly for vacancies and self-interstitials, the experimental knowledge is far from allowing to establish such dependences, and one pragmatically uses Fermi-level-independent values.

The situation becomes more complicated in case of dopants and other impurities which preferentially reside on substitutional sites as indicated in Fig. 35.7f. In such a configuration, the respective atoms are assumed to be immobile. To diffuse, they need to come into an interstitial configuration as depicted schematically in Fig. 35.7e, h or in a configuration with a vacancy as indicated in Fig. 35.7g. Diffusion as interstitials as in Fig. 35.7e is expected particularly for transition metals like gold and platinum. For boron in silicon, as another example, diffusion as pair with self-interstitials in analogy to Fig. 35.7h was suggested by Windl et al. [44] on the basis of ab initio calculations. Other elements like antimony in silicon and most dopants in germanium are assumed to diffuse via bound pairs with vacancies as shown schematically in Fig. 35.7g. Since even multiple exchanges of sites between vacancy and dopant would not lead to an effective movement, this mechanism requires that the vacancy partially dissociates from the dopant and returns to another nearest neighboring site of the impurity. In general, dopants will diffuse via both, self-interstitials and vacancies, although one of these mechanisms may dominate. Despite the atomistically completely different mechanisms, migrating interstitial atoms and pairs with intrinsic point defects are assumed to diffuse macroscopically as entities. As such, diffusion coefficients and charge states can be assigned to them, and their movement can be described in analogy to the concepts discussed above.

With substitutional impurity atoms being assumed immobile, the mobile dopant complexes must form dynamically. For boron and self-interstitials, De Salvador et al. [45] were able to clarify experimentally the complex reaction paths to mobile pairs. However, this is a rare case and for most other pair-formation reactions, such a detailed knowledge is not available. Accordingly, pair formation is usually described in a more generic way. Such pair-diffusion models were pioneered by Yoshida et al. [46] who proposed that substitutional impurities M_s form mobile MV pairs, depicted schematically in Fig. 35.7g, with vacancies V by the quasi-chemical reaction:



$$M_s + V \underset{k_{MV\leftarrow}}{\overset{k_{MV\rightarrow}}{\rightleftharpoons}} \mathrm{MV}. \tag{35.47}$$

The analogy for self-interstitials I is the formation of mobile MI pairs, shown schematically Fig. 35.7h, via

$$M_s + I \underset{k_{MI\leftarrow}}{\overset{k_{MI\rightarrow}}{\rightleftharpoons}} \mathrm{MI}. \tag{35.48}$$

Within the level of description of process simulation, it is not important whether MI is a bound impurity-self-interstitial pair or an interstitial impurity M_i. Interactions of self-interstitials with dopants were probably first considered by Seeger and Chik [47] in the form of their interstitialcy mechanism. For metal interstitials, one usually refers to the kick-out mechanism suggested by Gösele et al. [48]. In addition to the pairing reactions, one needs to include the bulk-recombination reaction

$$I + V \underset{k_{IV\leftarrow}}{\overset{k_{IV\rightarrow}}{\rightleftharpoons}} 0 \tag{35.49}$$

of self-interstitials and vacancies to an undisturbed lattice site symbolized by the "0." For metal diffusion, instead of the formation of vacancy pairs, one needs to consider the reaction of metal impurities with vacancies

$$M_i + V \rightleftharpoons M_s \tag{35.50}$$

proposed first by Frank and Turnbull [49] for copper in germanium. For the diffusion of dopants in semiconductors, such a reaction, just as the counterpart of a dopant-vacancy

pair reacting with a self-interstitial

$$\mathrm{MV} + I \rightleftharpoons M_\mathrm{s}, \tag{35.51}$$

constitutes a dopant-concentration-dependent parallel path to the recombination of self-interstitials and vacancies which also influences the concentrations of the pairs in steady state.

A quantitative description of the full system requires that the interactions between intrinsic point defects, substitutional impurities, mobile impurities and mobile complexes of impurities with intrinsic point defects, etc., are taken into account. This is usually done via the methodology of diffusion-reaction equations by writing, e.g., the interaction of species A with species B to species C in the form of the binary quasi-chemical reaction

$$A + B \underset{k_\leftarrow}{\overset{k_\rightarrow}{\rightleftharpoons}} C \tag{35.52}$$

with $k_\rightarrow$ and $k_\leftarrow$ being the forward and backward reaction constants. The reaction term

$$R = k_\rightarrow C_\mathrm{A} C_\mathrm{B} - k_\leftarrow C_\mathrm{C} \tag{35.53}$$

is common to the continuity equations of the species A to C

$$\frac{\partial C_\mathrm{A}}{\partial t} = \mathrm{div}\,(-J_\mathrm{A}) - R; \quad \frac{\partial C_\mathrm{B}}{\partial t} = \mathrm{div}\,(-J_\mathrm{B}) - R; \quad \frac{\partial C_\mathrm{C}}{\partial t} = \mathrm{div}\,(-J_\mathrm{C}) + R \tag{35.54}$$

and couples them. The different signs of the reaction term in the continuity equations mean that the concentrations of A and B are reduced when A and B react, while the concentration of species C increases. In the common case of several parallel reactions, the respective reaction terms are added or subtracted in the continuity equations of the species. For immobile species like substitutional dopants, the diffusion term is simply omitted in the continuity equation. In equilibrium or stead state, the concentration of the product can be expressed as

$$C_\mathrm{C}^\mathrm{eq} = C_\mathrm{A}^\mathrm{eq} C_\mathrm{B}^\mathrm{eq} \frac{\theta_\mathrm{C}}{\theta_\mathrm{A}\theta_\mathrm{B}} \exp\left(\frac{G_\mathrm{A}^\mathrm{f} + G_\mathrm{B}^\mathrm{f} - G_\mathrm{C}^\mathrm{f}}{kT}\right) = \frac{\theta_\mathrm{C}}{\theta_\mathrm{A}\theta_\mathrm{B}} \exp\left(\frac{G^\mathrm{B}}{kT}\right) \tag{35.55}$$

in terms of the respective concentrations of the reactants, the numbers of internal degree of freedom of the defects at specific sites θ_X, and their formation energies G_X^f. As long as charge is conserved in the reaction, the binding energy G^B will be independent of the Fermi level. A particular advantage of binary reactions is that the forward reaction rates can be estimated on the basis of Waite's theory of diffusion-limited reactions [50] in the form

$$k_\rightarrow = 4\pi a_\mathrm{R}\,(D_\mathrm{A} + D_\mathrm{B}) \tag{35.56}$$

as a function of the diffusion coefficients of the reacting species and the recombination radius a_R which is usually assumed to be on the order of the bond length between two lattice atoms. For the cases that both particles are electrically charged and that reaction barriers have to be overcome, extensions are available from Debye [51] and Waite [52], respectively. To determine the backward rate, two procedures have been used in literature. The first one is based on the concept that the backward reaction constant $k_\leftarrow$ is basically the break-up frequency of the defect. This is very similar to (35.43) with S^m and H^m replaced by the dissociation entropy and enthalpy of the defect C. The second method uses the fact that the forward and backward reaction rates of all reactions need to cancel out in equilibrium. Accordingly, the ratio of forward and backward reaction rate can be written in terms of the equilibrium concentrations of the species

$$\frac{k_\rightarrow}{k_\leftarrow} = \frac{C_\mathrm{C}^\mathrm{eq}}{C_\mathrm{A}^\mathrm{eq} C_\mathrm{B}^\mathrm{eq}} \tag{35.57}$$

which in turn can be written via (35.55) in terms of the binding energy of the reaction.

35.3.3 Macroscopic Diffusion Behavior of Dopants

Within the methodology of diffusion-reaction equations outlined in the previous subsection, mobile complexes are postulated to form via the reactions (35.47) and (35.48). In addition, the complementary reactions of the mobile complexes with intrinsic point defects via the reactions (35.50) and (35.51) need to be taken into account. To calculate the generation rate of the mobile complexes, concurrent reactions between substitutional dopants and intrinsic point defects in all charge states have to be included. For the mobile complexes, the single negative, neutral, and single positive charge states are usually considered, and steady state between them is assumed to be established rapidly. The resulting system of coupled partial differential equations then consists of the continuity equations for the self-interstitials and vacancies as well as for each dopant in continuity equations for the substitutional dopants and the pairs with self-interstitials and vacancies. Accordingly, this type of diffusion model is commonly referred to as "five-stream model." Although already a simplification, five-stream models are able to capture fairly complex diffusion phenomena. To illustrate some of

them, useful simplifications and their areas of application are discussed below.

As indicated first by Cowern et al. [53], a dynamic description of the formation of mobile dopant species is important to understand the macroscopic diffusion behavior for short annealing times. To illustrate this, we consider a simple system in which a substitutional impurity M_s changes with a frequency $1/\tau$ into the mobile state MI. Denoting the steady-state ratio of the concentrations in the substitutional and the mobile state by α and assuming that the diffusion flux follows Fick's first law (35.37), one arrives at a coupled system of equations in the form

$$\begin{aligned}\frac{\partial C_{M_s}}{\partial t} &= -\frac{1}{\tau}C_{M_s} + \frac{\alpha}{\tau}C_{\mathrm{MI}},\\ \frac{\partial C_{\mathrm{MI}}}{\partial t} &= \mathrm{div}\left(D_{\mathrm{MI}}\,\mathrm{grad}\,C_{\mathrm{MI}}\right) + \frac{1}{\tau}C_{M_s} - \frac{\alpha}{\tau}C_{\mathrm{MI}}.\end{aligned} \tag{35.58}$$

Assuming that all atoms are initially in a narrow layer in a depth of 0.5 µm, the total concentration $C_{\mathrm{M}} = C_{M_s} + C_{\mathrm{MI}}$ for $\alpha = 1000$ and D_{MI} chosen so that the macroscopic projected diffusion length for the time τ equals $\sqrt{2D_{\mathrm{MI}}\tau/\alpha} = 10$ nm is shown in Fig. 35.8 (coupled system) together with a simulation in which steady state between the substitutional and interstitial species is assumed (one equation). For the latter case, the diffusion equation results for all times in the Gaussian profiles shown. The same profiles are obtained for the coupled system for long diffusion times for which substitutional and mobile impurities are in steady state. In contrast, for the shortest time, the delta-doped layer is still clearly visible, and the profile has a characteristic shape with exponentially decreasing flanks.

For sufficiently long diffusion times, the substitutional impurities will be in steady state with the respective mobile species. To calculate concentrations of the mobile complexes, as emphasized by Cowern [54], one needs to consider the pairing reactions (35.47) and (35.48) as well as the reactions of the pairs with the opposite intrinsic point defects via the reactions (35.50) and (35.51). This leads in non-equilibrium situations to a nonlinear dependence of the pair concentration on both the self-interstitial and vacancy concentrations. However, for the important case that bulk recombination maintains local equilibrium between the intrinsic point defects ($C_{\mathrm{I}}C_{\mathrm{V}} \sim C_{\mathrm{I}}^{\mathrm{eq}}C_{\mathrm{V}}^{\mathrm{eq}}$), the same pair concentrations are predicted as in the case that the reactions (35.50) and (35.51) are ignored and the pair concentrations can be calculated from (35.55). Considering the diffusion of dopants, one has to keep in mind that acceptors on substitutional sites are negatively charged and donors positively. Accordingly, the concentrations of single negative, neutral, and single positive pairs C_{MX}, with X standing either for self-interstitials I or vacancies V, will be

$$\begin{aligned}&C_{\mathrm{MX}^-} = \eta_{\mathrm{MX}^-}C_{M_s^-}\frac{C_{\mathrm{X}}}{C_{\mathrm{X}}^{\mathrm{eq}}};\; C_{MX^0} = \eta_{\mathrm{MX}^0}C_{M_s^-}\frac{C_{\mathrm{X}}}{C_{\mathrm{X}}^{\mathrm{eq}}}\frac{p}{n_{\mathrm{i}}};\\ &C_{\mathrm{MX}^+} = \eta_{\mathrm{MX}^+}C_{M_s^-}\frac{C_{\mathrm{X}}}{C_{\mathrm{X}}^{\mathrm{eq}}}\left(\frac{p}{n_{\mathrm{i}}}\right)^2\end{aligned} \tag{35.59}$$

for acceptors and

$$\begin{aligned}&C_{\mathrm{MX}^-} = \eta_{\mathrm{MX}^-}C_{M_s^+}\frac{C_{\mathrm{X}}}{C_{\mathrm{X}}^{\mathrm{eq}}}\left(\frac{n}{n_{\mathrm{i}}}\right)^2;\; C_{\mathrm{MX}^0} = \eta_{\mathrm{MX}^0}C_{M_s^+}\frac{C_{\mathrm{X}}}{C_{\mathrm{X}}^{\mathrm{eq}}}\frac{n}{n_{\mathrm{i}}};\\ &C_{\mathrm{MX}^+} = \eta_{\mathrm{MX}^+}C_{M_s^+}\frac{C_{\mathrm{X}}}{C_{\mathrm{X}}^{\mathrm{eq}}}\end{aligned} \tag{35.60}$$

for donors. Therein, the $\eta_{\mathrm{MX}^{-/0/+}}$ stand for the temperature-dependent relative concentrations of pairs under intrinsic conditions ($n = p = n_{\mathrm{i}}$) with the respective point defects being in equilibrium ($C_X = C_{\mathrm{X}}^{\mathrm{eq}}$).

It should be noted that all the pair concentrations are proportional to the concentration of the substitutional atoms. In systems not too far from equilibrium, the concentrations of the pairs are expected to make up only a small fraction of the total concentration. As long as the concentrations of pairs remain small in comparison to the total dopant concentration, they will increase linearly with the respective point-defect concentration. In situations far from equilibrium, the formation of pairs will lead to a reduction of the substitutional concentration and consequently to a sublinear increase of the concentration of the dominating pair with the respective point defect concentration. In the limiting case, the redistribution of the dopants atoms takes place in the form of the dominating pairs with the respective diffusion coefficient.

In many important cases, the concentrations of pairs can be safely ignored. Then, the concentration of substitutional dopant atoms C_{M_s} in (35.59) and (35.60) can be replaced by the total dopant concentration C_M. Adding the continuity equations leads to the diffusion equation

$$\begin{aligned}\frac{\partial C_{\mathrm{M}}}{\partial t} = \mathrm{div}\Bigg(&D_{\mathrm{MI}}\,\mathrm{grad}\left(C_{\mathrm{M}}\frac{C_{\mathrm{I}}}{C_{\mathrm{I}}^{\mathrm{eq}}}\right) + D_{\mathrm{MV}}\,\mathrm{grad}\left(C_{\mathrm{M}}\frac{C_{\mathrm{V}}}{C_{\mathrm{V}}^{\mathrm{eq}}}\right)\\ &-z_{\mathrm{M}}\left(D_{\mathrm{MI}}\frac{C_{\mathrm{I}}}{C_{\mathrm{I}}^{\mathrm{eq}}} + D_{\mathrm{MV}}\frac{C_{\mathrm{V}}}{C_{\mathrm{V}}^{\mathrm{eq}}}\right)C_{\mathrm{M}}\,\mathrm{grad}\left(\frac{\psi}{U_{\mathrm{T}}}\right)\Bigg)\end{aligned} \tag{35.61}$$

with the diffusion coefficients D_{MX} given by

$$D_{\mathrm{MX}} = \eta_{\mathrm{MX}^-}D_{\mathrm{MX}^-} + \frac{\eta_{\mathrm{MX}^0}D_{\mathrm{MX}^0}p}{n_{\mathrm{i}}} + \eta_{\mathrm{MX}^+}D_{\mathrm{MX}^+}\left(\frac{p}{n_{\mathrm{i}}}\right)^2 \tag{35.62}$$

for pairs with acceptors and

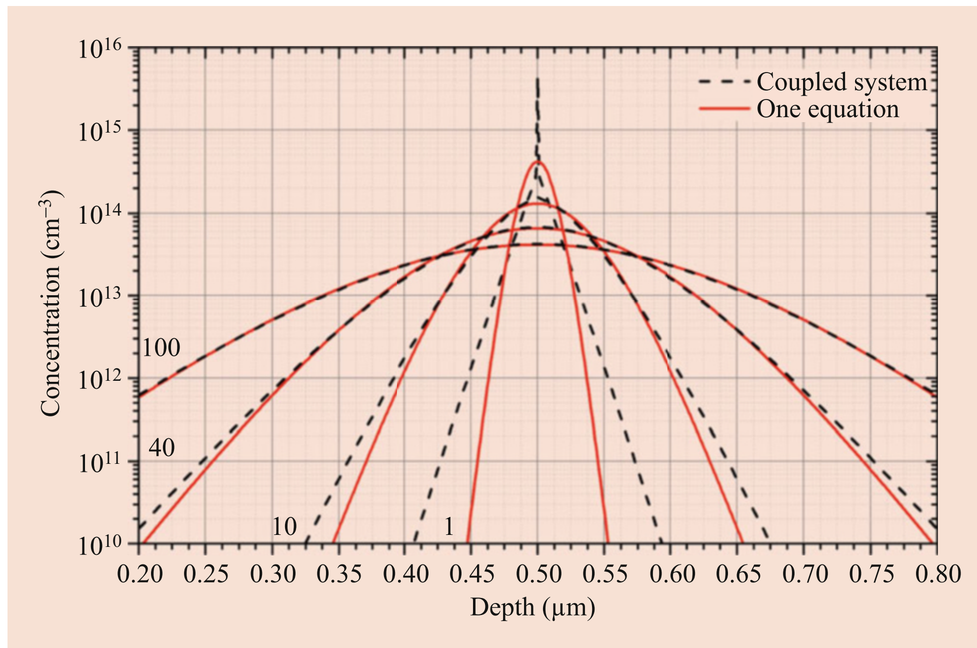


Fig. 35.8 Diffusion of impurities from an initially delta-doped layer, simulated by diffusion equations for mobile species and substitutional species and under the assumption of local equilibrium between these species. The running index is the diffusion time in multiples of the time constant τ. (After [5])

$$D_{\mathrm{MX}} = \eta_{\mathrm{MX}^-} D_{\mathrm{MX}^-} \left(\frac{n}{n_\mathrm{i}}\right)^2 + \eta_{\mathrm{MX}^0} D_{\mathrm{MX}^0} \frac{n}{n_\mathrm{i}} + \eta_{\mathrm{MX}^+} D_{\mathrm{MX}^+} \tag{35.63}$$

for pairs with donors. For both, the dependence of the diffusion coefficient on the Fermi level reflects the increase of the concentration of pairs in extrinsically doped semiconductors.

As long as the semiconductor is intrinsically doped, with vacancy and self-interstitial concentrations at equilibrium, and negligible strain effects, the respective diffusion coefficients of the dopants depend just on temperature. This "intrinsic diffusion coefficient" $D_0 = D_{\mathrm{MI}} + D_{\mathrm{MV}}$ shown in Fig. 35.9 for the main dopants in silicon and germanium is probably the best-studied property of dopants in semiconductors. An exception is boron in germanium for which only three experimental determinations are available which resulted in values which are orders of magnitude apart [55]. Comparing the other dopants for silicon and germanium, it is also noticeable that the fastest and slowest diffusers exchange places.

A further important property of dopants is the degree to which they diffuse via self-interstitials and vacancies. This is typically expressed in terms of the fractional diffusivity via self-interstitials:

$$f_\mathrm{I} = \frac{D_{\mathrm{MI}}}{D_{\mathrm{MI}} + D_{\mathrm{MV}}}. \tag{35.64}$$

The degree to which dopant atoms diffuse via vacancies and interstitial atoms was a subject of sharp discussions after Seeger and Chik [47] introduced the interstitialcy mechanism which lasted well after the review of Fahey et al. The current understanding of dopant diffusion in silicon was developed on the basis of diffusion studies in which some dopant atoms show enhanced diffusion, while the diffusion of others is retarded. Such studies, discussed in more detail in Sect. 35.3.5, in combination with information about the growth of extrinsic stacking faults led to the conclusion that boron and aluminum diffuse in silicon nearly exclusively via self-interstitials, while antimony has been found to diffuse nearly exclusively via vacancies. Arsenic in silicon diffuses via both, and phosphorus diffuses at low concentrations nearly exclusively via self-interstitials while at high concentrations a dominant diffusion via vacancies has been found. In germanium, all elements except boron were suggested to diffuse via a vacancy mechanism. For boron, the considerably lower diffusion coefficient in comparison to the other dopants is sometimes associated with an interstitial diffusion although no unambiguous experimental evidence is available.

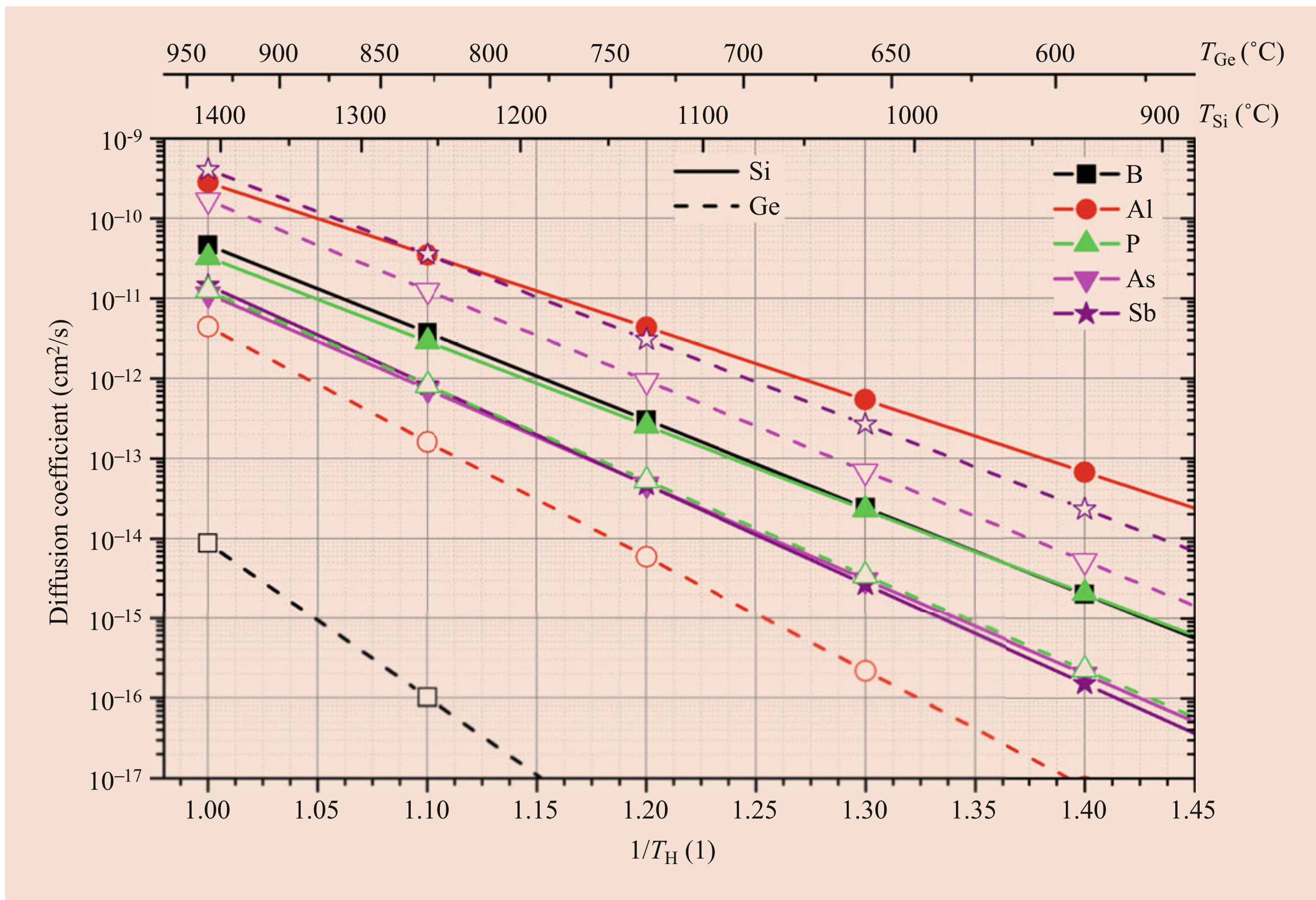


Fig. 35.9 Intrinsic diffusion coefficients of dopants in germanium and silicon as a function of the inverse homologous temperature. (The data are from [5] and [41] for silicon and germanium, respectively)

35.3.4 Dopant Diffusion at High Concentrations

Particularly for contact areas, the active doping concentration cannot be high enough. For such conditions, the charge carrier concentrations exceed the concentration n_i by far. The effects of such high doping levels on dopant diffusion will be discussed briefly in this subsection on the basis of an experiment reported by Orr Arienzo et al. [56] in which a highly doped polysilicon layer maintained a constant concentration of boron during the diffusion at 950 °C for 1 h.

The first simulation result shown in Fig. 35.10, denoted "Intrinsic diffusion," was obtained under the assumption of intrinsic diffusion with a constant diffusion coefficient. The resulting error function profile apparently underestimates the dopant penetration by far.

The second simulation result, denoted "+ Field effect," takes into account that the extrinsic doping concentration results in an electric field E which acts via (35.44) as an additional driving force for dopant redistribution. This manifests itself, as shown, e.g., by Smits [57], in an effective diffusion coefficient which depends via

$$D = D_0 \left(1 + \frac{C}{\sqrt{C^2 + 4n_i^2}}\right) \tag{35.65}$$

on the dopant concentration. However, the enhancement by a factor of two for $C \gg n_i$ remains too small to explain the experimental profile.

Considering also that boron diffuses predominantly via neutral pairs with self-interstitials, i.e., with an effective diffusion coefficient (35.62) which increases linearly with the hole concentration, one obtains the profile denoted "$D(p)$" in Fig. 35.10 under the assumption that the intrinsic point defects remain in equilibrium. In comparison to the previous simulations, the concentration-dependent diffusion coefficient results in pronounced platykurtic profiles which are flatter at the top but steeper at the flanks. Assuming an effective diffusion coefficient which increases quadratically with the respective charge carrier concentration would lead to a deeper penetration of the dopants but also to even steeper flanks.

To reproduce the gentler slope of the experimental profile as in the profile denoted "Full model," one has to take the

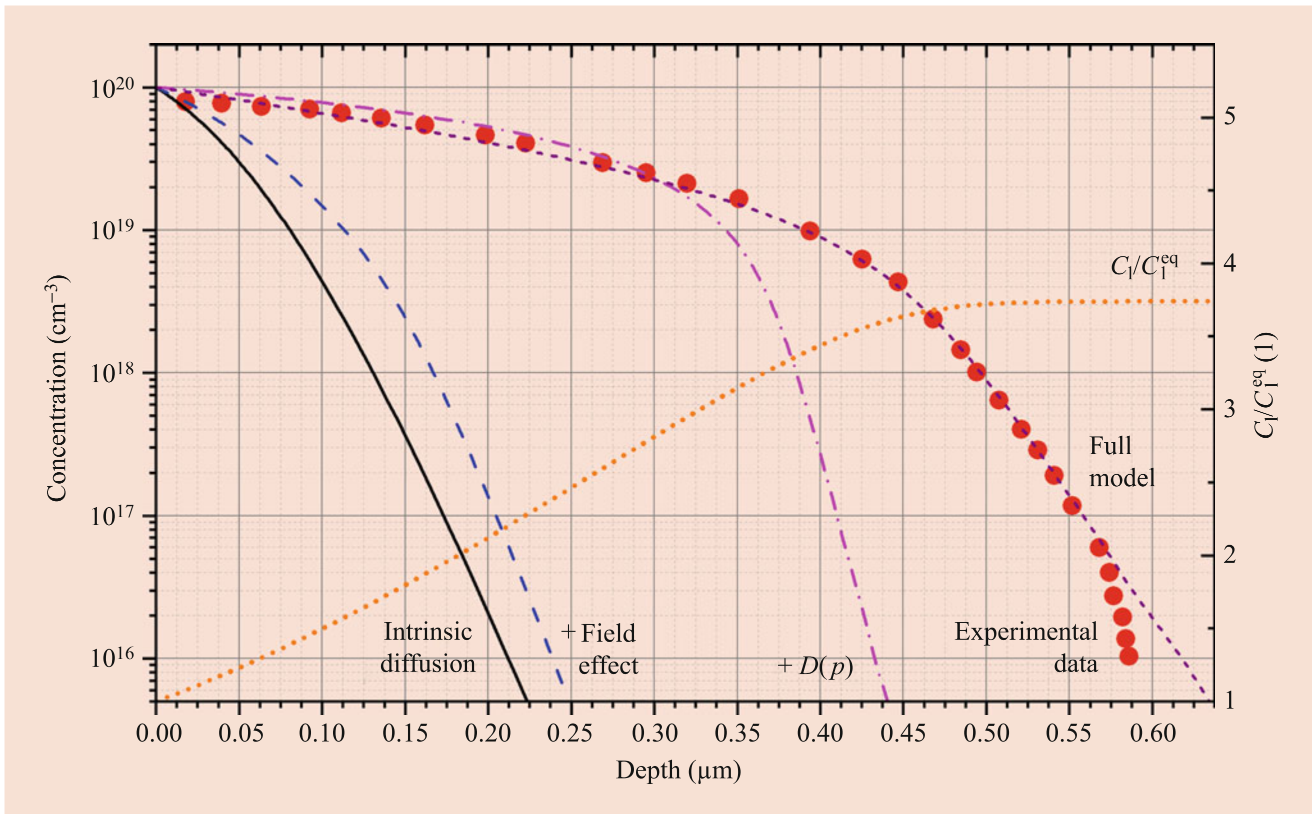


Fig. 35.10 Simulation of the diffusion of boron at high concentrations. (Courtesy C. Ortiz, Fraunhofer IISB) and comparison to the experiment of Orr Arienzo et al. [56] at 950 °C. After [5])

full interaction between boron atoms and self-interstitials into account: Boron-self-interstitial pairs form via (35.48) predominantly close to the interface to the polysilicon layer where the boron concentration is highest. When diffusing into the bulk, the boron concentration reduces, and the pairs are more and more likely to dissociate into substitutional boron and self-interstitials. This results, as shown by the line "C_I/C_I^{eq}" in Fig. 35.10, in an oversaturation of self-interstitials which increases with depth. There, the increase of the self-interstitial concentration compensates for the decrease of the effective diffusion coefficient with the charge carrier concentration so that the resulting doping profile becomes nearly error-function-like but with a much higher effective diffusion coefficient than in the case of intrinsic diffusion. Following their gradient, the self-interstitials will diffuse back to the surface so that their oversaturation depends on the interplay of transport into the bulk via dopant diffusion and transport back to the surface by self-diffusion. While the effect is rather unspectacular for boron, it is responsible also for the formation of pronounced kink and tail features in high-concentration phosphorus profiles and the broadening of the base region of bipolar transistors in areas below high-concentration phosphorus profiles, the so-called emitter-push effect [58].

At even higher phosphorus background concentrations, Nylandsted Larsen et al. [59] found that the diffusion of a variety of dopants increases steeply with the phosphorus concentration. Following the suggestion of Mathiot and Pfister [60], this was explained in terms of a percolation phenomenon. However, this explanation was discussed controversially, and Ramamoorthy and Pantelides [61], as an example, argued that the rapid diffusion should give rise to fast dopant clustering and the break-up of the percolation cluster.

35.3.5 The Influence of Surface Processes on Dopant Diffusion

First indications for the influence of surface chemical processes on dopant diffusion date back to the 1960s, when the (111)-oriented wafers prevalently used in industry were replaced by (100)-oriented wafers. Comparative diffusion studies at that time indicated an orientation-dependence of the diffusion of dopants which would not be expected for semiconductors with a diamond lattice. However, some researchers reported already then that such effects would be seen only for diffusion process in oxidizing ambient and not in inert ambient [62]. Later studies confirmed the oxidation-

enhanced diffusion of boron [63] and phosphorus [64] during dry oxidation of (100)-oriented silicon, while at the same time an oxidation-retarded diffusion of antimony was found [65]. Based on the observation of the concurrent growth of stacking faults identified unambiguously as self-interstitial agglomerates, it was suggested by Dobson [66] and Hu [67] that overstoichiometric silicon atoms are incorporated during the oxidation into the growing oxide layer and that these silicon atoms may segregate as self-interstitials into the silicon where they increase the concentration of self-interstitials beyond their equilibrium concentration. Because of bulk recombination of self-interstitials and vacancies (35.49), an oversaturation of the former ($C_\mathrm{I}/C_\mathrm{I}^\mathrm{eq} > 1$) will inevitably lead to an undersaturation of the latter ($C_\mathrm{V}/C_\mathrm{V}^\mathrm{eq} < 1$) and vice versa. The resulting effective diffusion coefficient under non-equilibrium conditions

$$D^\mathrm{eff} = D^\mathrm{eq}\left(f_\mathrm{I}\frac{C_\mathrm{I}}{C_\mathrm{I}^\mathrm{eq}} + (1-f_\mathrm{I})\frac{C_\mathrm{V}}{C_\mathrm{V}^\mathrm{eq}}\right) \qquad (35.66)$$

will be higher than the diffusion coefficient D^eq under equilibrium conditions ($C_\mathrm{I} = C_\mathrm{I}^\mathrm{eq}, C_\mathrm{V} = C_\mathrm{V}^\mathrm{eq}$) for dopants diffusing preferentially via self-interstitials ($f_\mathrm{I} > 0.5$). Dopants diffusing preferentially via vacancies ($f_\mathrm{I} < 0.5$), on the other hand, may be retarded in their redistribution. In process simulation programs, the injection of self-interstitials is usually modeled empirically as a flux of self-interstitials which increases sublinearly with the oxidation rate.

While the dry oxidation of (100)-oriented silicon was always found to result in an enhanced self-interstitial concentration, opposite effects were found for the diffusion of (111)-oriented samples at high temperatures and long diffusion times. This was explained by Taniguchi et al. [68] by a preferential segregation of self-interstitials into the growing oxide for small oxidation rates.

More problematic to explain was the retarded diffusion of boron and phosphorus, the enhanced diffusion of antimony, and the shrinkage of self-interstitial agglomerates reported by Mizuo et al. [69] during the nitridation of silicon. Interpreted in terms of (35.66), the nitridation-enhanced diffusion of antimony indicated an injection of vacancies. However, kinetic models in analogy to those for the injection of self-interstitials during oxidation had the problem that the growth of surface nitrides almost stopped after a short time, while the effects on dopant diffusion and stacking-fault growth lasted for much longer. In addition, it was shown by Ahn et al. [70] that similar effects can be observed below deposited nitrides. An explanation for these phenomena was finally suggested by Cowern [71] who noted that the work done against overlayers has to be taken into account in the formulation of the formation energies of the intrinsic point defects.

Finally, it should be noted that also other surface chemical processes like the nitridation of oxide layers or the formation of silicides were reported to lead to non-equilibrium diffusion effects.

35.3.6 Transient Diffusion Effects During Post-Implantation Annealing

Because of its superior homogeneity and dose control, ion implantation has become the main process for the doping of semiconductors. Inevitably associated with ion implantation is the creation of damage due to elastic collisions of the impinging atoms with the lattice atoms. Depending on the mass and energy of the implanted atoms, some ten to many hundred self-interstitials and vacancies are generated per implanted atom. For high enough doses, an amorphous layer is formed in the semiconductor. To remove the damage, semiconductors need to be annealed after ion implantation. With all the intrinsic point defects being far from equilibrium, such annealings result in significant implantation-enhanced diffusion of the dopants.

To illustrate typical effects during post-implantation annealing, simulations of a 20 keV boron implantation with a dose of 5×10^{15} cm^{-2} and subsequent isothermal annealings at 900 °C are shown in Fig. 35.11. During the first 10 to 40 s, significant diffusion is observed in the tail of the boron profile. During this time interval, diffusive broadening is limited to concentrations up to about 1×10^{19} cm^{-3}. This value corresponds to the concentration of electrically active, i.e., substitutional boron atoms, which remains nearly the same. Increasing the annealing time to 1 min results already in a much smaller broadening, and increasing the time further to 10 min has an even much smaller effect. In contrast, a significant increase of the electrically active boron concentration is observed. Because of their eminent technological importance, transient diffusion and transient activation have been main research topics for decades. For the sake of brevity, we can only present the most important results here and must refer the interested reader to some more extensive reviews in this area [73, 74].

While it was clear already at an early stage that implantation damage was the cause of the implantation-enhanced diffusion, the exact mechanism was not as clear for some time. The consensus today is that the implantation leads to an excess of self-interstitials in the crystal which agglomerate during annealing and undergo Ostwald ripening. The oversaturation which drives dopant diffusion results from the cloud of self-interstitials that the extended defects maintain in their vicinity [75]. Quantitative information about the extended defects could be obtained by numerous TEM (transmission electron microscopy) studies with the work of Eaglesham et al. [76] and Bonafos et al. [77] just to give some prominent examples. Complementary information about the energetics of self-interstitial clusters which are too small to be investi-

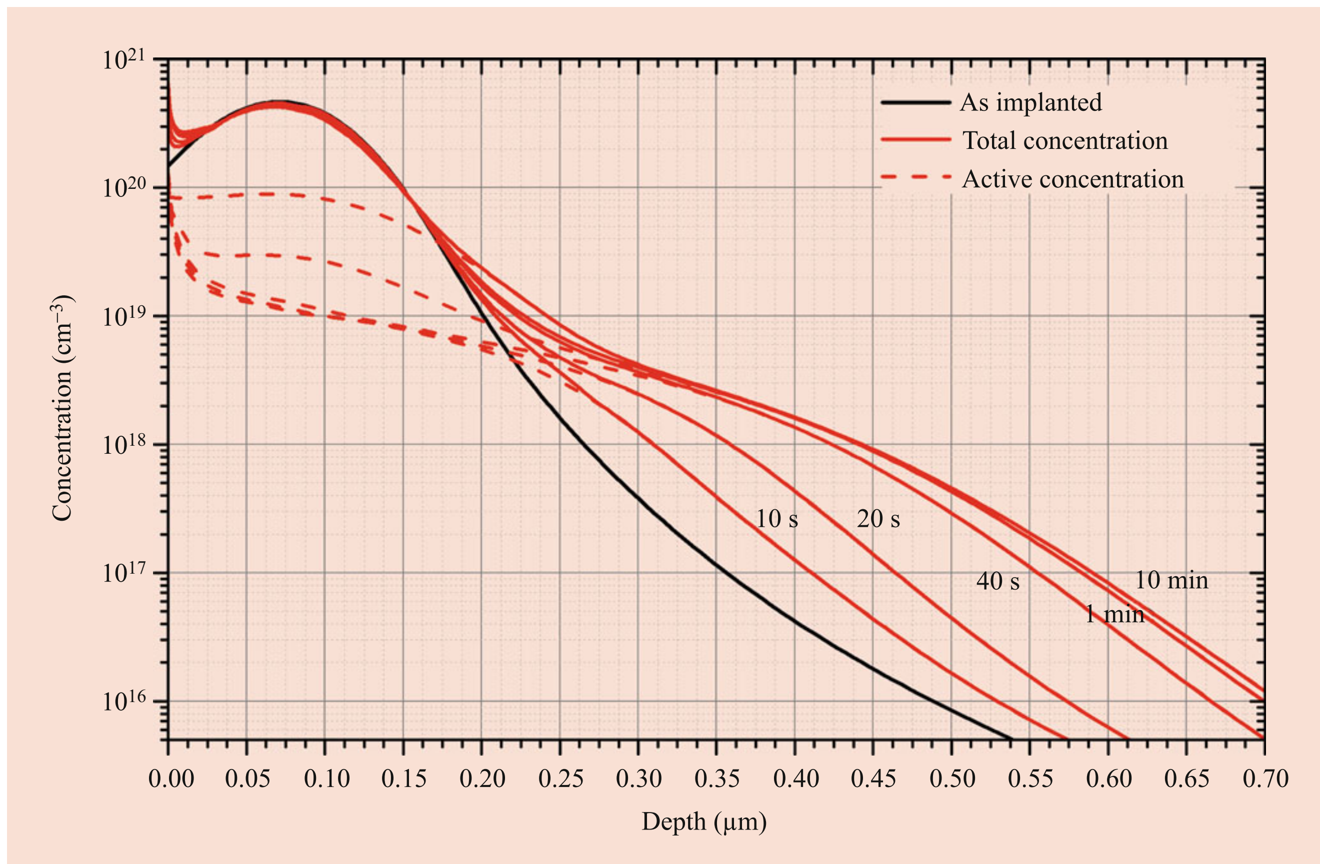


Fig. 35.11 Boron profiles after implantation with 20 keV and 5×10^{15} cm^{-2} and annealing at 900 °C simulated with Sentaurus Process [72]

gated by TEM became available by the experiment of Cowern et al. [78] in which the implantation-enhanced diffusion of narrow boron layers was induced by silicon implants. For a comprehensive review of the Ostwald ripening of extended defects, the review article of Claverie et al. [79] can be recommended. Among the many models suggested to capture the main aspects of the Ostwald ripening of small silicon clusters and extended {311} defects, the one of Zechner et al. [80] is particularly frequently used in process simulation since it combines predictive power with a minimum number of equations to be solved. For longer annealing times for which the {311} defects transform into stacking faults and perfect dislocation loops, extensions have been suggested by Zographos et al. [81] and Wolf et al. [82].

As initial conditions for the abovementioned models, the relevant concentration of self-interstitials needs to be calculated. This can, in principle, be done by the Monte Carlo techniques discussed in Sect. 35.2.2. However, the concentration of defects obtained would be far too high since vacancies and self-interstitials are created in pairs during the collisions with the impinging atoms and are likely to recombine early in the annealing stage. Accordingly, Giles [83] suggested for sub-amorphizing implants that only the extra atom introduced by the implantation needs to be taken into account as self-interstitial. While this assumption works well, e.g., for boron, it was found by Pelaz et al. [84] that the number of self-interstitials per implanted atom needs to be increased for heavier ions. Finally, for amorphizing implants, the amorphized region will regrow defect-free already at low temperatures like during the ramping stage of annealing. Accordingly, only the excess of self-interstitials in depths beyond the amorphous/crystalline interface needs to be taken into account and can be calculated by Monte Carlo simulations or by assuming an effective plus factor as for non-amorphizing implants [85].

35.3.7 Electrical Activation, Clusters, and Solid Solubility

One of the phenomena apparent from Fig. 35.11 is that a significant part of the atoms are immobile as well as electrically inactive. To explain such phenomena for arsenic in silicon, small electrically neutral clusters of two [86] and four arsenic atoms [87] were assumed to form. While the formation of such clusters can be described by binary reactions or chains of binary reactions, clusters Cl^n comprising n dopant atoms are usually assumed to form by the reaction

$$n\,M_s \underset{k_{\mathrm{Cl}\leftarrow}}{\overset{k_{\mathrm{Cl}\rightarrow}}{\rightleftharpoons}} \mathrm{Cl}^n \tag{35.67}$$

which results in an additional continuity equation

$$\frac{\partial C_{\mathrm{Cl}^n}}{\partial t} = k_{\mathrm{Cl}\rightarrow}\left(C_{M_s}\right)^n - k_{\mathrm{Cl}\leftarrow}C_{\mathrm{Cl}^n} \tag{35.68}$$

for the clusters. Because of mass conservation, the right-hand term multiplied by n also appears in the continuity equation of the substitutional impurities. Since the total concentration of impurities C_{M} includes now also the atoms in the cluster, adding all contributions leads again to an equation similar to (35.61) but with the substitutional concentration C_{M_s} instead of the total concentration C_{M} in the term on the right hand side. The substitutional concentration can now be considerably smaller than the total concentration, which leads already to a significant retardation of the diffusion of the impurities. In addition, the clustering also reduces the charge-carrier concentration in the high-concentration regions and thus also the effective diffusivity (35.62) of the non-clustered dopant atoms.

Coming back to the post-implantation annealing of boron shown in Fig. 35.11, it is noticeable that the electrical active concentration remains virtually the same during the period of transient-enhanced diffusion, while it increases significantly thereafter. This indicates an involvement of self-interstitials in the formation of the complexes responsible. Accordingly, for the latter, the term boron-interstitial clusters (BICs) was introduced in literature. Based on ab initio calculations, schemes of binary reactions of substitutional boron atoms and BICs with boron interstitial atoms and self-interstitials were developed (see, e.g., [88]) and shown to be able to capture the main features of transient boron diffusion and activation. For arsenic and phosphorus, the reaction schemes were even extended toward mixed clusters involving self-interstitials or vacancies [89].

Particularly by ion implantation, the total concentrations of dopants can be increased almost unlimitedly. However, when a certain total impurity concentration, known as solid solubility, is exceeded, an impurity-rich phase is expected to form in equilibrium. Depending on the impurity and temperature, the impurity-rich phase may be a liquid, a pure impurity phase, or a silicide phase [90]. Dynamically, such a phase can form by spinodal decomposition or, often more likely, by nucleation and growth of precipitates. Since the latter process is kinetically limited, dopant precipitates are often observed only after annealing processes with high thermal budgets [91]. However, dopant precipitates are not only observed after ion implantation. During growth and annealing of phosphosilicate glasses, as a prominent example, so many phosphorus atoms segregate into the silicon that the (binary) solid solubility of phosphorus in silicon is exceeded there and SiP precipitates form [92].

35.3.8 Segregation

In real applications, semiconductors will always be in contact to other materials like silicon dioxide, silicides, or metals. During annealing processes, the atoms being present in either material will diffuse not only therein but also across the interfaces between the materials. In equilibrium, because of their different formation energies, discontinuous distributions of the atoms are expected at the interfaces. An example in this sense already mentioned above is the formation of large SiP precipitates in silicon where the phosphorus concentration is 50 at.% in the SiP, while the total concentration in silicon corresponds largely to the temperature-dependent solid solubility. At the interfaces, the discontinuous distributions of atoms in equilibrium are characterized by the segregation coefficients m as ratios of the respective concentrations at the two sides. During thermal processing, the flux of species between the phases A and B is usually written in form of a first-order kinetic model

$$J_{\mathrm{X}} = h\left(C_{\mathrm{X}}^{\mathrm{A}} - mC_{\mathrm{X}}^{\mathrm{B}}\right) \tag{35.69}$$

as being proportional to the interface transport coefficient h and the difference in concentration normalized by the segregation coefficient.

Segregation of dopants from highly doped phosphosilicate or borosilicate glasses into silicon is used in solar and power electronics for the doping of silicon with concentrations which may be even above solid solubility. At lower concentrations, diffusion of dopants in silicon dioxide is rather slow so that segregation effects can be observed only when the interface moves during an oxidation process. During such processes, phosphorus, arsenic, and antimony pile up at the silicon side, while boron segregates preferentially into the growing oxide.

Pile-up of donors was observed also at stationary interfaces between silicon and silicon dioxide. Dedicated investigations showed that this pile-up occurs for phosphorus [93] and arsenic [94, 95] within few nm from the interface on the silicon side even if the dopants in the area below are homogeneously distributed. In contrast, a similar pile-up of boron was shown to result rather from large self-interstitial gradients toward the interface during post-implantation annealing [96]. In continuum simulators, interface segregation of phosphorus and arsenic is usually included via an approach suggested by Lau, Orlowski, and coworkers [97, 98]. It assumes that dopants adsorb and desorb from both materials to energetically favorable sites at the interface.

35.3.9 Simulation Methodologies

The dawn of numeric diffusion simulation began around 1980 with programs like SUPREM [6, 7] and ICECREM [8], when diffusion theories were already too complicated to be solved analytically. These programs like their two- and three-dimensional successors were based on a discretization of the diffusion-reaction equations discussed above. A big advantage of such continuum approaches is that the mesh used for discretization of the equations can be adjusted to the problem, i.e., made fine where necessary, while it can be left coarse elsewhere. A second potential advantage is that empirical models like the clustering Eq. (35.67) or the self-interstitial-complex ripening models mentioned in Sect. 35.3.6 can be implemented easily to reduce parameters and equations to be solved. On the other hand, there is an over-linear increase of computer resources and simulation time associated with an increase of the number of equations to be solved. Due to their excellent prediction accuracy and robustness for a wide range of applications from nanoelectronics to power electronics, continuum models are currently the workhorses of TCAD simulations in industry.

As an alternative, atomistic simulators like DADOS [99] were introduced at the end of the 1990s. These programs consider directly the diffusive jumps of non-lattice atoms and defects as well as the reactions among them by a kinetic Monte Carlo (KMC) algorithm. The big advantage of KMC simulators is that a plethora of different defects and defect reactions can be taken into account with a linear increase of computational efforts only. On the other hand, all non-lattice atoms and defects must be taken along in the simulations which make them feasible only for comparatively small volumes. However, with the continuing miniaturization of devices, this is becoming all the easier.

To simulate crystal-orientation-dependent phenomena like faceting during epitaxial deposition or shape changes during laser annealing, both continuum and KMC approaches lack predictive power. For such effects, it is necessary to include the host atoms in the calculations as in lattice kinetic Monte Carlo (LKMC) schemes [100]. However, it should be noted that classical LKMC implementations lack the ability to simulate other defects apart from vacancies and substitutional impurities.

35.4 Simulation of Oxidation

For many years, the simulation of oxidation was both a very important and a very difficult part of process simulation: Local oxidation ("LOCOS") was indispensable to generate three-dimensional structures for electrical isolation. It led to transition domains between the thin gate oxide and the isolation oxide, as shown in Fig. 35.12. The basic effect is that during oxidation 2.27 volumes of oxide are generated from one volume of silicon. For local oxidation as shown in Fig. 35.12, the excess volume of oxide generated leads to a flux of oxide and the need to solve the resulting mechanical problem.

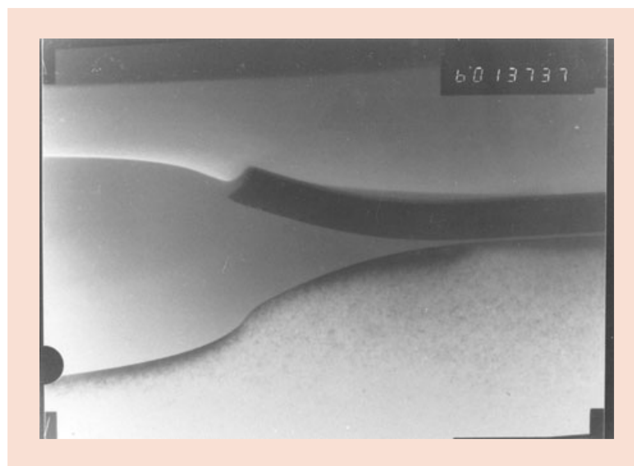


Fig. 35.12 SEM micrograph of transition domain ("bird's beak") between thick field oxide and thin gate oxide. Technology from the 1980s, with gate oxide thickness of about 100 nm

Simulation of oxidation includes simultaneously the diffusion of dopants and defects in the bulk semiconductor and in the (growing) oxide, the mechanics of the oxide flux, and the transport of dopant atoms across the moving interface between the semiconductor and the oxide, which in many cases critically affected the total amount of dopants near the channel, and in turn even top electrical parameters like the threshold voltage. In turn, the generation and adaptation of meshes to properly resolve the transport of dopants across moving interfaces was a key problem for three-dimensional process simulation.

For current nanoelectronic devices, the simulation of oxidation is far less important, because advanced device architectures do not include LOCOS oxidation, but merely thin oxides. Shallow trench oxidation is performed by growth of a thin oxide followed by oxide deposition. However, for applications like power electronics, thicker oxides are still needed, and the three-dimensional simulation of oxidation stays important for such cases.

35.4.1 One-Dimensional Simulation of Oxidation

The basic physical processes to be included in the one-dimensional simulation of oxidation are shown in Fig. 35.13. They are all fast compared with the oxide growth and can therefore be considered as stationary. At the interface between oxide and silicon, the oxidizing species O_2 in case of

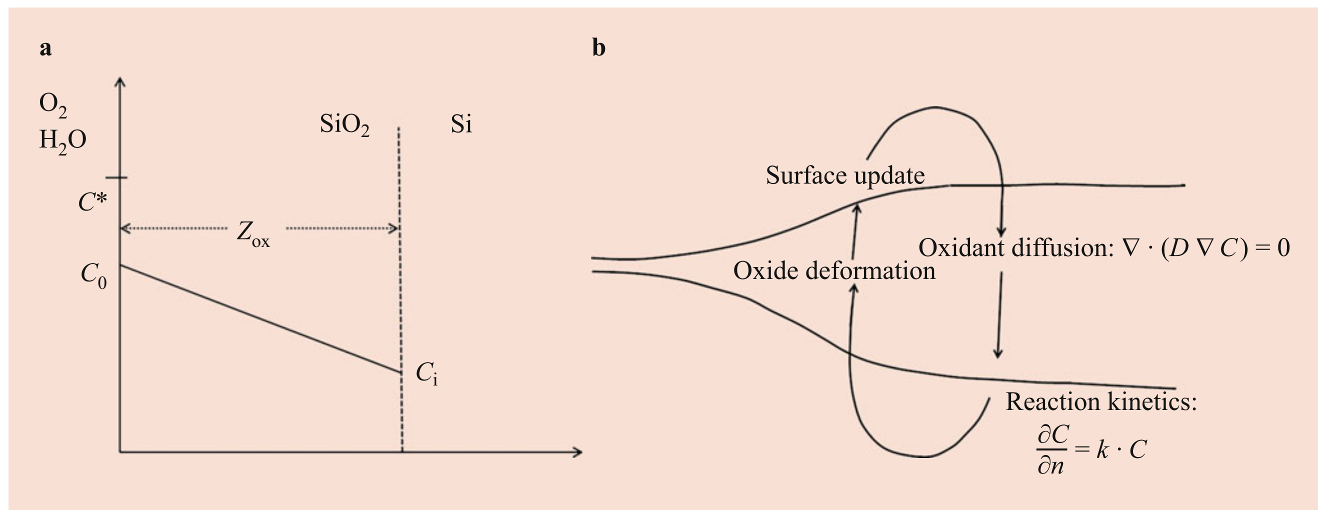
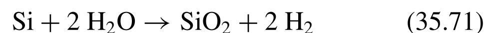


Fig. 35.13 Physical processes involved in the oxidation of silicon. (**a**) One-dimensional case (planar oxidation); (**b**) Two−/three-dimensional case as discussed in Sect. 35.4.2

dry oxidation and H_2O in case of wet oxidation react with silicon according to

$$\mathrm{Si} + \mathrm{O_2} \rightarrow \mathrm{SiO_2} \quad (35.70)$$



$$\mathrm{Si} + 2\,\mathrm{H_2O} \rightarrow \mathrm{SiO_2} + 2\,\mathrm{H_2} \quad (35.71)$$

The corresponding Deal-Grove model [101] consists of three steps, which can all be expressed as fluxes F of the oxidizing species. These species are first adsorbed at the oxide surface, resulting in a concentration C^*, which depends on the partial pressure of the oxidizing ambient (O_2 or H_2O). A parameter h describes the transfer of the species across the surface, giving rise to the concentration C_0 at the oxide side of the surface:

$$F = h \cdot \left(C^* - C_0\right) \quad (35.72)$$

Stationary diffusion with a diffusion coefficient D leads to a concentration C_i of the oxidizing species at the oxide side of the interface to silicon:

$$F = D \cdot \frac{(C_0 - C_\mathrm{i})}{z_\mathrm{ox}} \quad (35.73)$$

A constant k describes the first-order reaction of the oxidizing species at the interface between oxide and silicon:

$$F = k \cdot C_\mathrm{i} \quad (35.74)$$

Since the three fluxes in Eqs. (35.72), (35.73), and (35.74) are equal, a differential equation for the oxide thickness z_ox results [101]:

$$\frac{\mathrm{d}z_\mathrm{ox}}{\mathrm{d}t} = \frac{kC^* / N}{1 + \frac{k}{h} + k \cdot z_\mathrm{ox} / D} \quad (35.75)$$

Here, N is the density of the atoms in the silicon crystal. The solution of Eq. (35.75) is the well-known linear-parabolic relationship for the time dependence of the oxide thickness:

$$\frac{z_\mathrm{ox}{}^2}{k_\mathrm{p}} + \frac{z_\mathrm{ox}}{k_\mathrm{l}} = t + t_0 \quad (35.76)$$

Here, k_p and k_l are the parabolic and the linear oxide growth coefficients. They relate to the physical parameters given above according to

$$k_\mathrm{l} = \frac{C^*}{N\left(\frac{1}{k} + \frac{1}{h}\right)} \quad (35.77)$$

$$k_\mathrm{p} = 2D \cdot C^* / N \quad (35.78)$$

The linear regime applies for small oxide thicknesses z_ox and is determined by the transfer of the oxidant across the oxide surface and the reaction at the interface between oxide and silicon, expressed by the parameters k and h. For larger oxide thickness, the transport through the existing oxide layer limits the process, resulting in a growth asymptotically

approaching the square root of the oxidation time. The initial time t_0 is used to express the native oxide which is grown almost instantaneously. An increase of the initial oxidation rate can be described by Massoud's model [102], which adds an exponential term to Eq. (35.75):

$$\frac{\mathrm{d}z_{ox}}{\mathrm{d}t} = \frac{kC^* / N}{1 + \frac{k}{h} + k \cdot z_{ox} / D} + C\,\mathrm{e}^{\left(-\frac{z_{ox}}{L}\right)} \tag{35.79}$$

In this case the oxide thickness is no more given in the analytical form of Eq. (35.76), but must be calculated as solution of the differential Eq. (35.79).

In Eqs. (35.75), (35.76), (35.77), (35.78), and (35.79), the physical parameters D, k, h, C, and L depend in various respects on the process to be simulated, including temperature, the orientation of the silicon crystal, the oxidizing species (O_2 or H_2O) and its partial pressure, additional HCl in the atmosphere, the dopant concentration on the silicon side of the interface between silicon and oxide, and stress.

35.4.2 Multidimensional Simulation of Oxidation

The one-dimensional oxidation model summarized above describes the situation where a homogeneous oxide layer is grown on a flat silicon surface. This does no more apply in areas where the silicon is structured (e.g., a trench) or partly masked. This situation is visualized in two dimensions in Fig. 35.13b. In this case, Eqs. (35.72), (35.73), and (35.74) are replaced by a stationary partial differential equation and the corresponding boundary conditions for the concentration C of the oxidizing species, depending on space and time. Here, $\frac{\partial}{\partial n}$ is the derivative in direction normal to the surface or interface:

$$\nabla\left(D \cdot \nabla C\right) = 0 \text{ in the oxide;} \tag{35.80}$$

$$D \cdot \frac{\partial C}{\partial n} = h \cdot \left(C^* - C\right) \text{ at the oxide surface;} \tag{35.81}$$

$$D \cdot \frac{\partial C}{\partial n} = k \cdot C \text{ at the interface between oxide and silicon.} \tag{35.82}$$

The dependence of the physical constants D, h, and k on the process to be simulated is in principle similar to the one-dimensional case. However, for non-planar or masked structures, the stress-dependent flux of the oxide must be simulated, which results from the generation of 2.27 volumes of oxide from one volume of silicon. In addition to the standard Arrhenius dependence on the temperature, this leads to

$$D = D_0 \cdot \exp\left(-\frac{E_D}{kT}\right) \cdot \exp\left(-\frac{p \cdot V_D}{kT}\right); \tag{35.83}$$

$$C^* = C_0 \cdot \exp\left(-\frac{E_C}{kT}\right) \cdot \exp\left(-\frac{p \cdot V_C}{kT}\right); \tag{35.84}$$

$$k = k_0 \cdot \exp\left(-\frac{E_k}{kT}\right) \cdot \exp\left(\frac{\sigma_{nn} \cdot V_k}{kT}\right); \tag{35.85}$$

$$\eta = \eta_0 \cdot \exp\left(-\frac{E_\eta}{kT}\right) \cdot \exp\left(p \cdot \alpha(T)\right) \tag{35.86}$$

where the symbols have partly been adapted from the literature [103]. Here, E_d, E_C, E_k, and E_η are the activation energies in the respective Arrhenius laws, with k the Boltzmann constant. p is the hydrostatic pressure and σ_{nn} the normal stress at the interface between oxide and silicon, counted as positive if the oxide executes a pull on the silicon. V_D, V_c, V_k, and V_η (see Eqs. (35.83), (35.84), (35.85), (35.86), and(35.87)) are the activation energies for the stress dependencies of the respective parameters and α a parameter depending on temperature. Whereas according to [103], V_k is the difference between the molecular volume of SiO_2 and the atomic volume of Si, and equals to 25 Å^3, the other volumes are fitting parameters. It was shown [104] that for the oxidation of cylindrical structures, Eq. (35.86) leads to an equation for the hydrodynamic pressure p, which has no solution in case of some concave structures. It was therefore suggested to replace Eq. (35.86) by a model proposed by Eyring [105, 106]:

$$\eta = \eta_0 \cdot \exp\left(-\frac{E_\eta}{kT}\right) \cdot \frac{\sigma / \sigma_c}{\sinh\left(\sigma / \sigma_c\right)}, \text{ with } \sigma_c = 2kT / V_\eta \tag{35.87}$$

35.4.3 Multidimensional Simulation of Stress-Dependent Oxidation

Besides the specific physical models for the diffusion and reaction of the oxidizing species also the mechanical problem resulting from the generation of 2.27 volumes of oxide from

one volume of silicon must be solved numerically. In general, this needs the simulation of the viscoeleastic oxide flow. For this, the stress $\boldsymbol{\sigma}$ and strain $\boldsymbol{\varepsilon}$ must be considered. They are given by tensors (here displayed as six-dimensional vectors), and the hydrodynamic pressure p relates to them as follows:

$$\boldsymbol{\sigma} = \left(\sigma_x, \sigma_y, \sigma_z, \tau_{xy}, \tau_{xz}, \tau_{yz}\right) \tag{35.88}$$

$$\boldsymbol{\varepsilon} = \left(\varepsilon_x, \varepsilon_y, \varepsilon_z, \gamma_{xy}, \gamma_{xz}, \gamma_{yz}\right) \tag{35.89}$$

$$P = -\left(\sigma_x + \sigma_y + \sigma_z\right)/3 \tag{35.90}$$

Here, the indices of the first three components give the direction of the stress or strain.

For the shear stresses in last three components of the stress $\boldsymbol{\sigma}$, the first index indicates the normal to the surface to which the shear stress is applied, and the second index gives the direction of the shear stress. In this nomenclature one could also use σ_{xx} instead of σ_x, etc. The stain components give the changes of the shifts u, v, and w in directions x, y, and z:

$$\varepsilon_x = \frac{\partial u}{\partial x}; \varepsilon_y = \frac{\partial v}{\partial y}; \varepsilon_y = \frac{\partial w}{\partial y}; \tag{35.91}$$

$$\gamma_{xy} = \frac{\partial u}{\partial y} + \frac{\partial v}{\partial x}; \gamma_{xz} = \frac{\partial u}{\partial z} + \frac{\partial w}{\partial x}; \gamma_{yz} = \frac{\partial v}{\partial z} + \frac{\partial w}{\partial y} \tag{35.92}$$

The space available here does not allow for the discussion of the various approaches and approximations used for the simulation of stress in the literature. It is important to note that for temperatures of 960 °C and above, the viscous behavior of the oxide dominates [107] whereas below the oxide increasingly becomes elastic. In the viscous case, the velocity field $\boldsymbol{V}$ of the oxide flow can then be related to the hydrodynamic pressure P via

$$\eta \cdot \Delta \mathbf{V} = \nabla P \tag{35.93}$$

If the oxide is approximated as incompressible, Eq. (35.93) reduces to $\Delta \mathbf{V} = 0$. In that case all stress dependencies given in Eqs. (35.83), (35.84), (35.85), and (35.86) are neglected and only that Laplace equation is used to simulate the flow of the incompressible oxide caused by the extra volume generated during oxidation. This also means that masks are considered as flexible, executing no pressure on the growing oxide.

Figure 35.14 shows a sample oxide structure with the different kinds of interfaces at which boundary conditions for

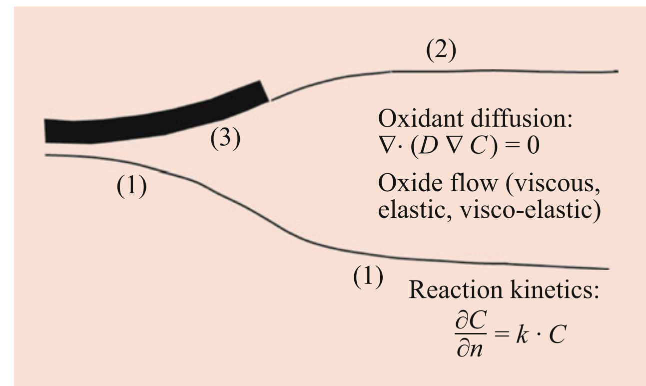


Fig. 35.14 Boundary conditions for multidimensional simulation of oxidation

$\boldsymbol{v}$ and P must be defined: (1) At the interface between oxide and silicon, $\boldsymbol{v}$ is normal to the interface, and its value is given by the rate at which silicon is transformed into oxide (equal to 0.44 times the oxide growth rate); (2) at the oxide surface, P is equal to the ambient pressure minus the surface tension; (3) at the interface between oxide and nitride, P is equal to the mechanical pressure executed by the nitride layer. This needs either mechanical calculations of the nitride bending or the use of suitable analytical approximations. For the then still missing boundary condition for the velocity at the free oxide surface, Chin used an additional relationship provided by an artificial compressibility algorithm [108].

35.5 Lithography Simulation

One of the most important process steps in semiconductor manufacturing is lithography, because it enables the transfer of the chip design onto the silicon wafer. Here, three properties are most important: the capability to generate tiny features (currently, the smallest dense features for single exposure are in the range of 20 nm); to establish the pre-defined links foreseen in the chip design between the transistors and capacitors acting as switches, amplifiers, or storage elements; and moreover to process up to hundreds of billion elements at the same time. Because a separate chapter of this handbook is dedicated to the lithography equipment and technology, below only a few technological aspects are outlined, which are necessary to present lithography simulation in a self-contained section.

35.5.1 Basic Principle of Lithography

Figure 35.15 shows the principle of a projection lithography step. Light from a suitable source illuminates a mask, where it is diffracted at the mask features. The diffracted light is

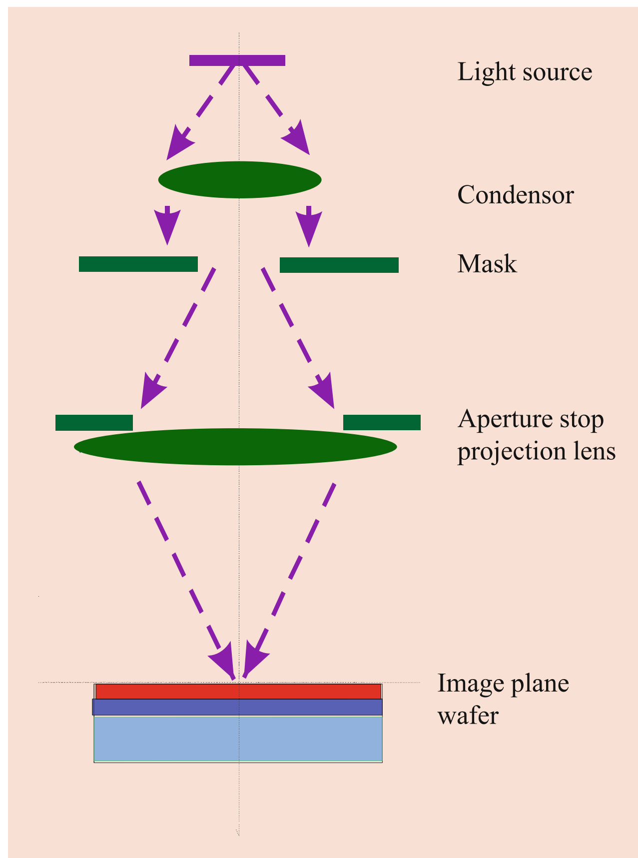


Fig. 35.15 Schematics of imaging in optical lithography

then collected via a suitable optical system and focused onto a photoresist. This part of the lithography process is called imaging and depends both on the mask and the lithography equipment, namely, the light source and the optical system. The resolution of this optical system is then given by the Rayleigh criterion [109].

$$d_{\min} = k_1 \cdot \frac{\lambda}{\mathrm{NA}} \tag{35.94}$$

where $d_{\min}$ is the minimum pitch which can be printed, λ is the wavelength of the light used, NA is the numerical aperture (equal to the relative refractive index of the medium times the sine of the opening angle of the last lens), and k_1 is the so-called technology factor, with a minimum value of 0.5. Because 193 nm is the smallest wavelength of laser light currently used in high-volume semiconductor device manufacturing, for which lenses are sufficiently transparent, the theoretical minimum pitch which can be printed in the so-called "dry" optical lithography (with air as medium between the last lens and the photoresist) is about 104 nm. Smaller pitches can either be printed via so-called multiple patterning lithography [110], with immersion lithography [111] or with extreme ultraviolet (EUV) lithography [112]. In the latter case, soft X-rays at a wavelength of about 13.5 nm are used, which then necessitate the replacement of the lenses in all parts of the optical system by multilayer MoSi mirrors. The projection process transfers an image of the mask patterns into the photoresist, leading to local deposition of energy into the photoresist. The image is diffraction limited, because the lenses cannot collect all diffraction orders from the mask. Furthermore, the image is disturbed by several other effects such as defocus, lens imperfections, or defects on the mask. After resist development parts of the resist which were exposed (positive tone resist) or not exposed (negative tone resist) are chemically modified and in turn (partly) removed. This latter part is influenced by chemistry and diffusion effects in the resist.

35.5.2 Principle of Lithography Simulation

Figure 35.16 illustrates the lithography process and its simulation, which is divided into two steps: First, for the imaging step, the intensity distribution, generated by the projection of the mask through the optical system, inside the photoresist is simulated. Second, the modification of the photoresist caused by the energy deposited during exposure, and governed by chemical reactions, diffusion, and temperature in case of a post-exposure bake process, is simulated. Different to the other process steps described in the other subsections, the imaging step requires simultaneous simulation on equipment, and feature-scale level to calculate from characteristics of the illumination source, the mask, and the optical system used the energy deposition in photoresists on a nanometer scale.

35.5.3 Simulation of Imaging

The physics of projection is well understood and could in principle be fully described by the well-known Maxwell's equations. However, due to the different size dimensions to be simulated during the projection step, different simulation methods derived from the Maxwell's equations must be applied. For example, the projection of the mask can be simulated very accurately with derivatives of the Hopkins's equation [113], whereas the mask diffraction is typically simulated by directly solving the Maxwell's equations either without any simplifications or, depending on the mask area and accuracy to be simulated, by using some derived simplified methods.

One common derivative of the Hopkins' equation is the Abbe approach [111], which basically consists of a Fourier transformation F of the mask transmission, a multiplication with pupil functions P in the Fourier space to describe the transmission behavior of the optical system, and finally an inverse Fourier transformation F^{-1} to simulate the intensity

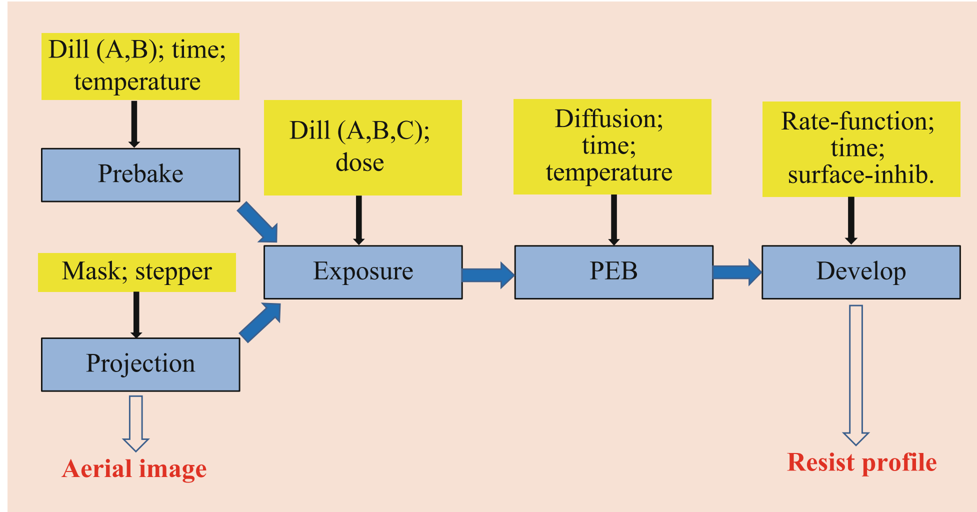


Fig. 35.16 Principle of lithography simulation

distribution in the image plane. This approach is based on the assumptions of an illumination with a single plane wave along the z-axis or with a tilt angle (t_x, t_y), a given complex mask transmission $t(x,y)$, and projection along the z-axis. Spatial frequencies f_x and f_y are introduced, which depend on the diffraction angles Θ_x and Θ_y in x- and y-directions, and the wavelength λ, and are used for the mask diffraction spectrum $s(f_x, f_y)$. In the following formulas, a scalar description not considering polarization is given:

$$f_x = \frac{\sin(\theta_x)}{\lambda}; f_y = \frac{\sin(\Theta_y)}{\lambda} \tag{35.95}$$

$$s(f_x, f_y) = F\{t(x,y)\} = \int t(x,y) \cdot \mathrm{e}^{-i2\pi\pi(f_x \cdot x + f_y \cdot y)} \mathrm{d}x\, \mathrm{d}y \tag{35.96}$$

For illumination with the tilt angle (τ_x, τ_y), the mask diffraction spectrum is shifted accordingly:

$$s(f_x, f_y) \rightarrow s\left(f_x - \frac{\sin\tau_x}{\lambda}, f_y - \frac{\sin\tau_y}{\lambda}\right) \tag{35.97}$$

The pupil function $P(f_x, f_y)$ expresses within the numerical aperture NA, for example, focus and aberrations, and is zero outside:

$$P(f_x, f_y) = 0 \text{ for } \sqrt{\sin^2\theta_x + \sin^2\theta_y} > \text{NA} \tag{35.98}$$

The complex amplitude a_P of the light in the image plane resulting from a single source point P is then given by the inverse Fourier transformation of the pupil function times the diffraction spectrum, where the tilt is considered in the pupil function P:

$$a_P(x,y) = F^{-1}\left\{P\left(f_x - \frac{\sin\tau_x}{\lambda}, f_y - \frac{\sin\tau_y}{\lambda}\right) \cdot F\{t(x,y)\}\right\} \tag{35.99}$$

Alternatively, the tilt could also be considered in the diffraction spectrum s according to Eq. (35.97) while using the pupil function P without tilt.

The overall light intensity is then calculated as the incoherent sum across the images of all points P of the light source:

$$I_{\text{tot}} = \sum_P a_P(x,y) \cdot a_P^*(x,y) \tag{35.100}$$

However, for state-of-the-art lithography technologies, the real topography of the masks and their materials, described by the complex refractive indices $n + ik$, must be considered as well as polarization effects. This is not possible with analytical approaches like the Kirchhoff approximation [114], which assumes an infinitely thin mask. Moreover, for some investigations, it is not sufficient to only calculate the image generated in the image plane (the so-called aerial image). Instead, the light transmission within the photoresist must be calculated, because the intensity changes due to absorption and reflection.

A possible approach, not using the Kirchhoff approximation, is to simulate the diffraction at the mask via an efficient numerical solution of Maxwell's equations, for example, via the so-called finite-difference time-domain (FDTD) method (see the standard reference [115]) or in Fourier space via a variant of the so-called rigorous coupled-wave analysis (RCWA) (see [116]). The key feature of FDTD is that Maxwell's equations are solved on a staggered grid, one subgrid for the electrical field $\boldsymbol{E}$ and another subgrid for the magnetic field $\boldsymbol{H}$, as visualized in Fig. 35.17: Here, the values of one field vector (either $\boldsymbol{E}$ or $\boldsymbol{H}$) at the time

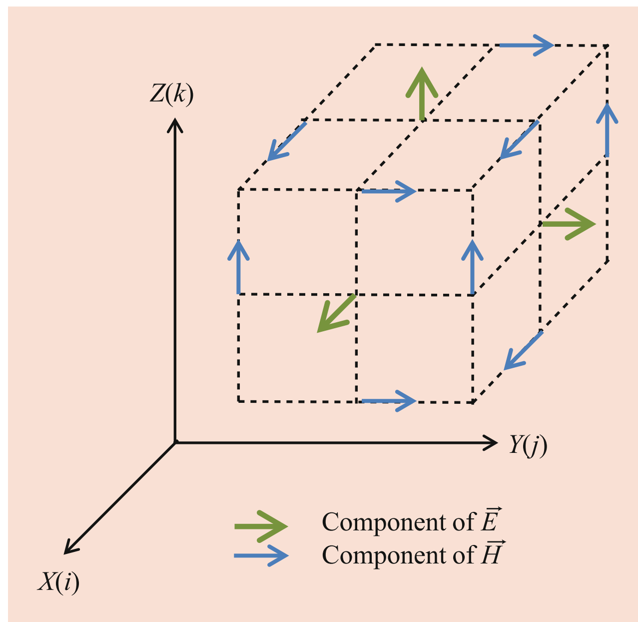


Fig. 35.17 Spatial discretization in the FDTD algorithm: distribution of E and H across the two staggered grids

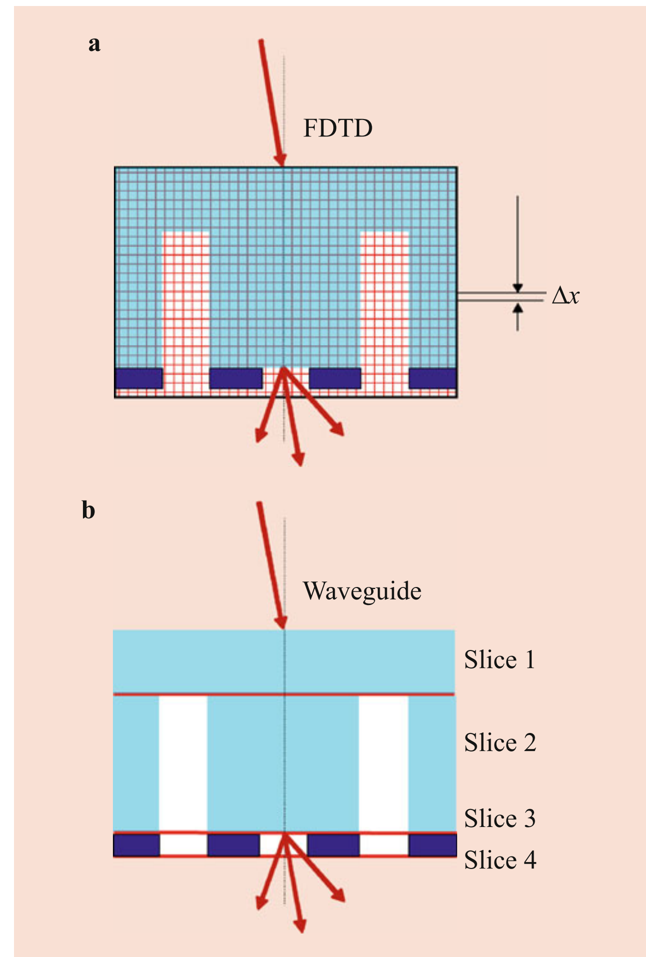


Fig. 35.18 Schematic mask representation for (**a**) FDTD and (**b**) waveguide. The figure shows one and a half periods of a standard alternating phase-shift mask AltPSM and the corresponding FDTD mesh and waveguide slices

t_n is calculated from the values of the other field vector (either $\boldsymbol{H}$ or $\boldsymbol{E}$) at the preceding time step t_{n-1}. Both the accuracy and the computation time are about proportional to the number of mesh points. Because the typical mesh size is about 5% of the wavelength λ, the computation time scales with about $1/\lambda^2$ [117]. The method is more appropriate for optical lithography than for EUV lithography where the ratio between wavelength and feature size is much smaller than in optical lithography. Figure 35.18a shows a schematic discretization of an alternating-phase-shift mask AltPSM.

The waveguide method (WG) [117] is a variant of the RCWA which is well suited for the simulation of the impact of EUV masks. Such masks consist of absorbers placed on top of the multilayer MoSi mirror. For the simulation the mask is divided into slices with homogeneous optical properties in the vertical direction; see Fig. 35.18b. Inside these slices or waveguides, both the electromagnetic fields and the permittivity profiles are expanded into Fourier series and are inserted into Maxwell's equations. The problem statement based on Maxwell's equations is transformed into the solution of an eigenvalue problem, which yields the eigenmodes of all mask slices. Using proper boundary conditions, the coupling of the modes of all slices leads to the overall physical response of the system. The computation time scales linearly with the number of inhomogeneous slices (more than one material inside the slice) and cubically with the numbers of the modes considered in both lateral directions. Both FDTD and WG and their advantages and drawbacks are explained more in detail in the literature [117].

Another approach used for the simulation of parts of resist exposure is the transfer matrix method for homogeneous layers. Here, the light intensity within the layer stack is calculated for a given number of layers with their refractive indices $n + ik$, angle of incidence, and polarization. The two variables considered are the amplitude of the incident and the reflected light within the system. Incident and reflected light within a layer propagate as a plain wave with damping constant k, described by a diagonal matrix. Transmission and reflection at the interfaces between layers depend on the complex refractive indices of the layer together with the angles of incidence and the polarization of the light and are expressed by a matrix where the off-diagonal elements describe the reflection. Besides the orientation of the mask features, the difference in the reflectivity between a transversal-electric and a transversal-magnetic wave may result in important polarization effects. A more detailed description of the transfer matrix method can be found in the literature, for example [114].

35.5.4 Simulation of Resist Development

The imaging process outlined above modifies the photoresist due to the locally deposited energy. In the subsequent development step, depending on the resist type used, either the (sufficiently) exposed or the (sufficiently) unexposed parts of the resists are removed.

The common description of this process is given by the Dill model [118]:

$$\frac{\partial M\,(r,t)}{\partial t} = -C\,\tilde{I}\,(r,t)\,M\,(r,t) \qquad (35.101)$$

$$a\,(r,t) = A\,M\,(r,t) + B \qquad (35.102)$$

Here, the concentration of the photoactive compound $M(r,t)$ decreases due to the deposited energy density $\tilde{I}\,(r,t)$, and the absorption $\alpha(r,t)$ depends on $M(r,t)$. A, B, and C are the Dill model parameters for the photoresist in question. Furthermore, the photoactive compound M diffused with a diffusion coefficient D:

$$\frac{\partial M\,(r,t)}{\partial t} = D\Delta M\,(r,t) \qquad (35.103)$$

Finally, the local resist development rate $R(M)$ depends on M. Various refinements of this standard reaction-diffusion model were developed in order to describe state-of-the-art photoresists, e.g., chemically amplified resists (CAR).

35.5.5 State-of-the-Art Lithography Simulation

Current state-of-the-art lithography simulation tools employ selections of the basic approaches summarized below in various combinations, depending on the tool and the application in question. Commercially supported tools such as PROLITH [119], Sentaurus-Lithography [120], and OPTOLITH [121] additionally provide elaborated user interfaces, and especially Sentaurus-Lithography and OPTOLITH are integrated with other TCAD tools of the respective vendors. Examples shown in the following subsection have been generated with the research and development simulator Dr.LiTHO [122], which is designed especially for efficiency and flexibility in the implementation of physical models and algorithms.

The semi-analytical Kirchhoff approximation allows for some simple simulation studies where the mask can be approximated as a two-dimensional pattern of transparent and opaque arrays and where polarization effects can be neglected. State-of-the-art masks need to be simulated with numerical approaches such as FDTD, RCWA, or WG, which are also mandatory if polarization comes into play. Whereas these are needed for the near field of the mask, the imaging of this near field onto the wafer may further be simulated, for example, with the Abbe approach outlined above, utilizing the result of the mask simulations. As described above, energy deposition in the photoresist is simulated with the transfer matrix approach, and resist development is being simulated based on the Dill model.

Besides these aspects which are specific to lithography simulation, several more general techniques are also needed, customized, and employed in lithography simulation: During resist development, its surface is moved with one of the techniques described below in Sect. 35.7 on deposition and etching. In order to further improve the efficiency of simulation and to extend the possible sizes and use cases of the masks to be simulated, different domain decomposition methods [123] can be employed which may include not only three-dimensional but partly also two- or even one-dimensional simulations. This may also reuse results of prior simulations which were stored in a repository [123]. To enable important and promising technological approaches such as source-mask co-optimization [124], advanced global optimization algorithms such as genetic algorithms [125] are being employed. Furthermore, as mentioned below in Sect. 35.8 on process variations, lithography simulation is also an essential part of TCAD-based Design-Technology Co-Optimization (DTCO) approaches, in order to establish the link between mask layout and the performance of real devices and circuits.

35.5.6 Examples for State-of-the-Art Lithography Simulation

In the following, a few examples of results obtained with the research and development lithography simulator Dr.LiTHO [122] in the context of current research on the development of advanced lithography technologies are shown.

Figure 35.19 shows the impact of a particle located on an EUV mask pellicle on the printed feature size (critical dimension, CD) for an EUV system with a NA of 0.55 under specific imaging conditions [125]. The pellicle is a polysilicon membrane at 2.5 mm mask distance. One can see that larger particles are critical even in the presence of a mask protection membrane.

Figure 35.20 shows the impact of mask absorber 3D effects on the process window of 20 nm semi isolated lines of a NA 0.33 EUV system [123]. The black curve shows the result for an ideal absorber in combination with a real multilayer and the gray curve the same for a real absorber in combination with a real multilayer. In the simulations, the real absorber can be replaced by an ideal one to "switch off" the absorber-related 3D mask effects. The comparison

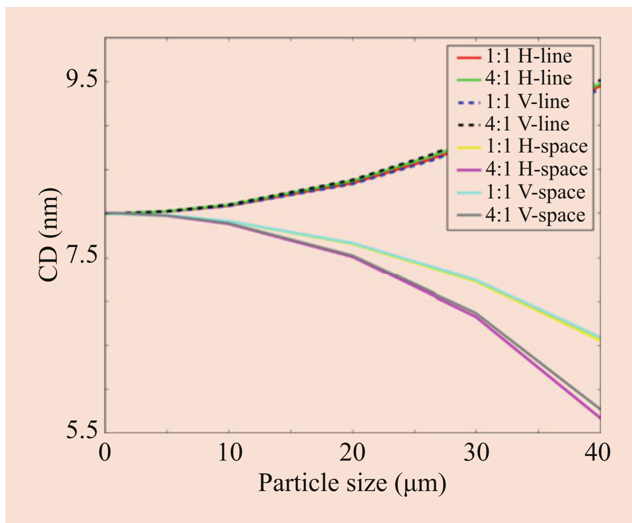


Fig. 35.19 Example for the impact of a particle located on an EUV mask pellicle on the printed CD, for different duty factors *n*:*m* of lines and spaces. The lines and spaces are located either in horizontal (H) direction (perpendicular to the chief ray angle of the system) or in vertical (V) direction (parallel to the chief ray angle of the system) [125]

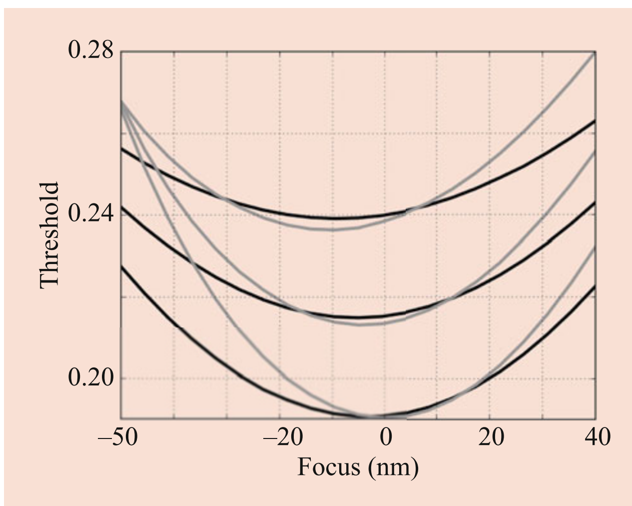


Fig. 35.20 Process windows of an ideal EUV absorber (black, infinitely thin absorber) and a real EUV absorber (gray) both on top of a real EUV multilayer [123]

with the fully real mask shows the effect caused by the real absorber geometry, which is mainly the expected asymmetric through focus behavior.

The defect-induced loss of reflected light and image intensity in the area of an EUV multilayer defect can be partly compensated by a removal of absorber in the vicinity of the defect. Figure 35.21 demonstrates a simulation of this method for 32 nm features at a NA of 0.33 [126]. The repair is done by removal of the absorber in a circle around the center of the defect; see dashed larger circles in the figure, with an optimum of 50 nm (wafer scale). Other defect positions and sizes may require more complex repair shapes, especially at larger NAs where more details of the defect will become visible.

35.6 Simulation of Deposition and Etching

Similar to lithography, also for deposition and etching, it is indispensable to take equipment effects into account when simulating the process results on feature scale. However, here mostly a clear separation can be made between the equipment and the feature-scale level: Equipment simulation (on a scale of centimeters to meters) can be used in a preparatory step to calculate quantities just above the (macroscopically flat) wafer, such as local gas or particle flows and local temperatures, which are affected by the reactor geometry, process recipes, and partly varying further parameters. These intermediate results can then be used as inputs for feature-scale simulation (on the scale of micro- to nanometers) of the local deposition and etching of layers. Figure 35.22 visualizes the separation between these simulation scales.

The area of the simulation of deposition and etching is very diverse in three respects: First, it includes macroscopic equipment and nanoscale feature simulation, and moreover also the (generally three-dimensional) simulation of the evolvement of partly complex surfaces. These aspects are reflected by the following subsections. Second, there is a huge variety of etching and deposition processes which cannot be covered within the space available for this section. Therefore, here only the basic principles and approaches are discussed, and no attempt is made to list the specific equations and parameters for the large variety of processes used in semiconductor technology.

35.6.1 Outline of Equipment Simulation

Whereas in the area of lithography integrated equipment and feature-scale process simulation is mandatory since long, as described in the preceding section, for both deposition and etching, the simulation at equipment level can be separated from the feature-scale simulation. Similar to this, this separation is also possible for diffusion and oxidation, whereas in these two areas, equipment simulation is rarely used. Except for plasma doping, e.g [127], for ion implantation equipment simulation has so far not been an issue.

In most cases equipment simulation consists of the following steps:

1. Three-dimensional discretization of the process reactor or furnace.
2. Establish assumptions on the physical/chemical modeling

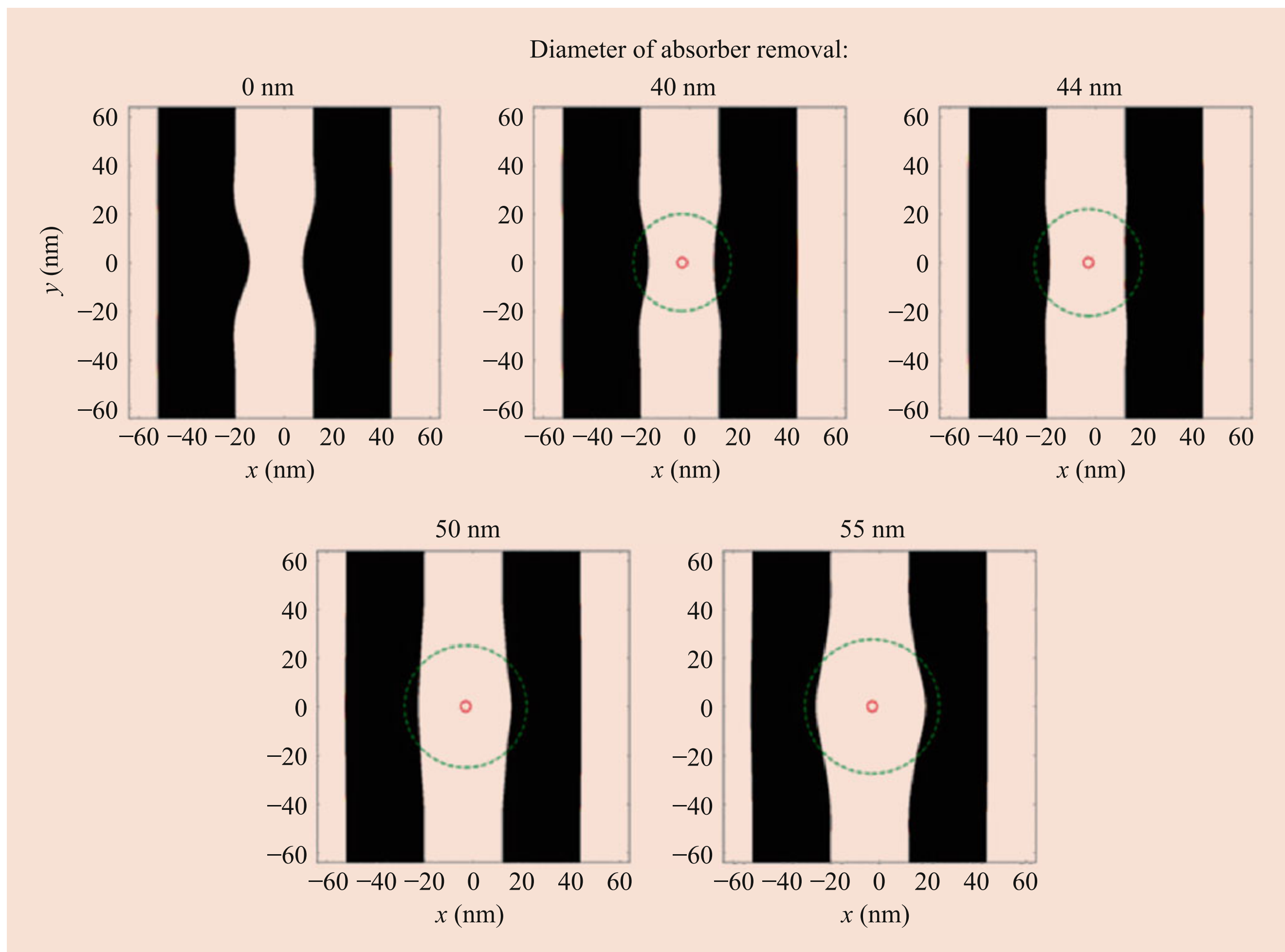


Fig. 35.21 EUV multilayer defect repair by increasing absorber removal around the defect. The defect caused line deformation is reduced with an optimum at 50 nm absorber removal [126]

components for the system in question, e.g., gas flow, resistive or inductive heating, radiative heat transfer, plasma physics, characteristics of sputter targets, chemical reactions, etc.

3. Set up and solve a system of partial differential equations, representing fluid dynamics, thermal behavior, plasma characteristics, electrical properties, and chemical reactions for the equipment, possibly including macroscopic models for the interaction with the wafers (e.g., as sinks or sources for heat or for chemical species). Such macroscopic models do not have to resolve the individual features on the wafer. This step can be omitted if, e.g., the particle flow can be extracted directly in step 4 from the characteristics of a sputter target defined in step 2.
4. Extract from the equipment simulations the quantities of interest directly above the wafer, e.g., temperature, pressure, species concentration, angular or energy distributions of ions, electrical field, or potential (without considering individual features on the wafer).

Whereas all these steps are necessary to define and describe the equipment and process in question, the main numerical effort is needed for step 3, except for cases where no fluid dynamics is involved. In order to efficiently implement step 3, usually well-established computational fluid dynamics [128] or plasma simulation codes [129] are employed, which then have to be supplied by the user with the relevant physical data and models, as described above.

As an example, Fig. 35.22a deals with the equipment simulation of the deposition of silicon oxide in a capacitively coupled plasma-enhanced chemical vapor deposition (PECVD) reactor using oxygen and TEOS chemistry. The contour colors in the figure represent the concentration of the oxygen radicals. The corresponding fluxes of oxygen radicals, together with the fluxes of oxygen ions, provide the boundary conditions for the feature-scale simulation, the result of which is shown in Fig. 35.22b. This figure represents a cross section of a contact hole with a non-conformally deposited silicon oxide layer.

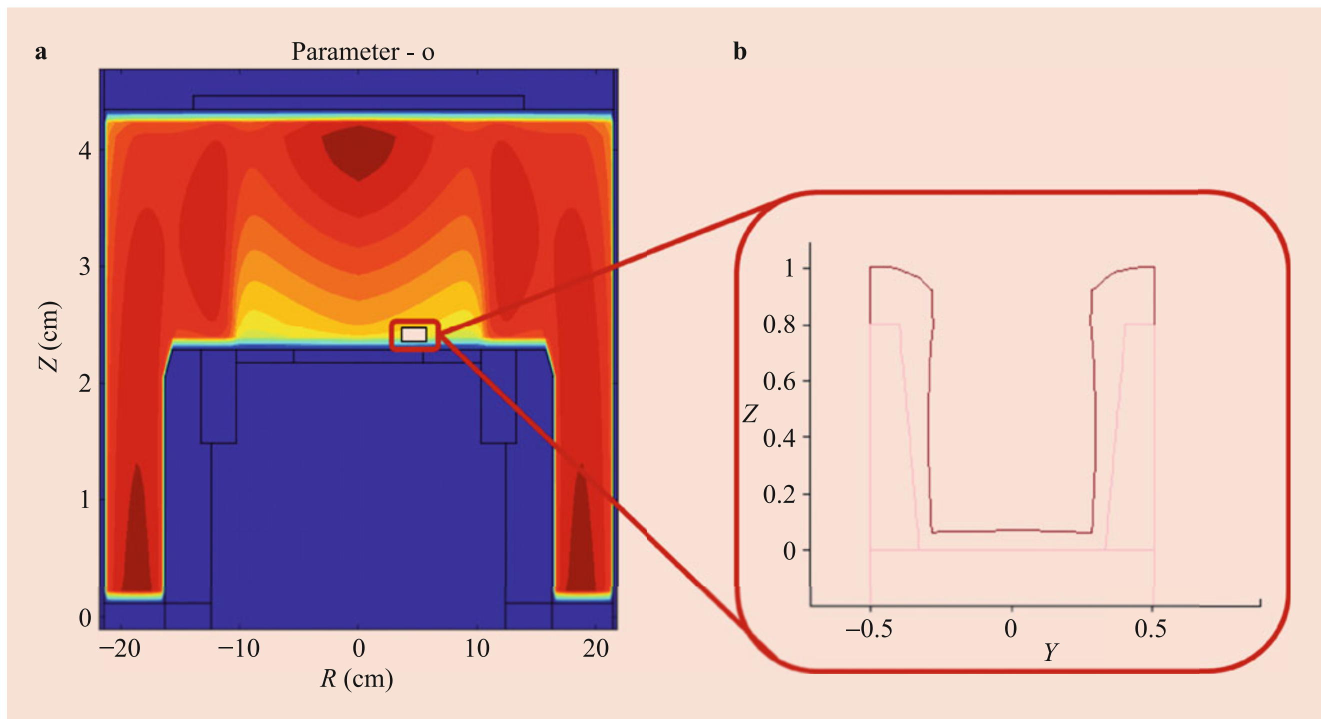


Fig. 35.22 Visualization of the scales of (**a**) equipment and (**b**) feature-scale simulation for deposition and etching. The clipping for feature-scale simulation in the left figure is not to scale

In case of diffusion and oxidation processes, the situation is somewhat easier, because the process gas does not react in the furnace volume, but only on the wafer (in case of oxidation, transforming silicon to silicon oxide). Here, step 3 is reduced to the simulation of fluid dynamics and thermal behavior. However, several transient effects might need to be considered, due to the movement of a wafer or a batch of wafers into or out of the preheated furnace or due to ramping the temperature up and down while the wafer is being processed. This is especially important in case of very short time annealing processes, such as spike or millisecond annealing, which need elaborated simulation of the transient local temperature distributions because these are not in equilibrium. For example, laser thermal annealing was discussed in [130].

35.6.2 Discretization and Movement of Surfaces

Deposition and etching do not only change the geometry of the device or structure during fabrication, but may also change its topology: New layers may be added, old layers may be completely removed, originally connected features may get disconnect and vice versa, and holes may be transferred into voids. Very small features may be critical, because they may, e.g., act as an etch stop. Also curved, tilted, and non-axis-aligned structures must be accurately described. In turn, it is vital to employ surface discretizations and algorithms for their update, which are numerically efficient and stable, can reliably handle arbitrary topographies and their changes and allow for adaptive control of the local number of elements used, to achieve a good compromise between accuracy and efficiency.

Three main methods have been used for the description of geometries and of their changes during processing. The simplest and most stable is the cell-based description and the related cell-removal algorithm [131], which is, however, due to its inherent limitations hardly used any more and here only mentioned for historical and reference purposes. The other two approaches are triangulation and the level set method. Triangulation is the three-dimensional extension [132] of the "string algorithm" which was generally used for two-dimensional topography simulation [133]. Here, surfaces and interfaces are discretized by triangles. Geometrical changes during processes are then described by moving the vertices of the triangles according to rates which result from physical models as described in Sect. 35.6.3. The main advantage is that these models can directly use the position, size, and orientation of these triangles. The main disadvantage is that it is difficult with this approach to deal with topological changes, e.g., in case that during layer deposition the opening of a hole closes to form a void, as it would happen if the process shown in Fig. 35.23 would be continued.

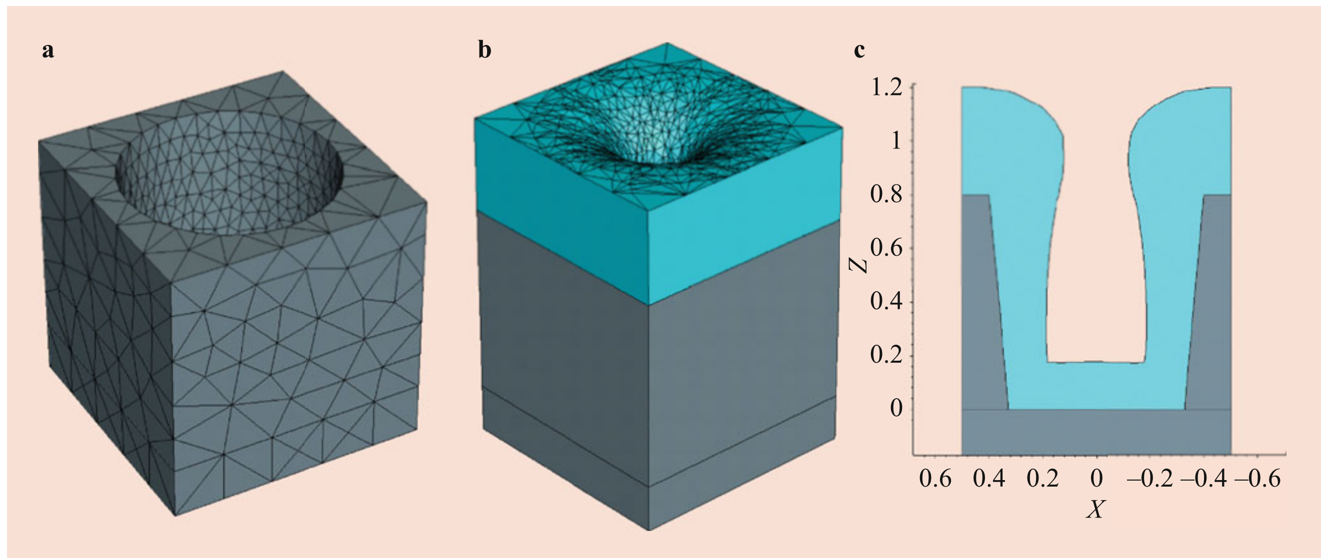


Fig. 35.23 Example for the discretization of a three-dimensional structure using triangles. (**a**) Initial contact hole; (**b**) simulation of nonconformal deposition of low-temperature oxide (LTO). Here, the side-wall faces of the simulation structure are discretized as rectangles; (**c**) cross-sectional view. The scale shown is in microns

The approach which is most frequently used for the discretization of three-dimensional surfaces and interfaces is the level set method: Here, the surface is described as the set of points $\boldsymbol{x}$ where a three-dimensional function $F(\boldsymbol{x}, t)$ equals zero, and the movement of the surface is simulated based on local rates $\boldsymbol{r}(\boldsymbol{x}, t)$:

$$\frac{\partial F(\boldsymbol{x}, t)}{\partial t} + \boldsymbol{r}(\boldsymbol{x}, t) \cdot \nabla F(\boldsymbol{x}, t) = 0 \quad (35.104)$$

The main advantages are that with this approach surfaces and interfaces are given as solutions of partial differential equations, which are solved on a three-dimensional mesh, similar to bulk variables like dopant concentrations, and that changes in topology should not lead to problems if the underlying three-dimensional grid is locally fine enough. The disadvantage is that surface elements and surface normals needed in the implementation of the physical models as outlined in Sect. 35.6.3 are not given directly and must be extracted separately from the level set description.

35.6.3 Models for Deposition and Etching Rates

Physical/chemical models for deposition and etching consider the interaction of different ionic or neutral species with the surface to allow the determination of local deposition or etching rates. Boundary conditions for the feature-scale modeling can be provided by equipment simulations or measurements, e.g., of temperature, pressure, species concentrations, or energies of neutrals and ions. In the following, as an example we discuss low-pressure chemical vapor deposition, where the mean free path is large compared with the feature size and the particles move isotropically in the volume of the reactor far away from the surface.

Assuming a triangulated surface, for each triangle I the reaction flux R_i depends on the flux of particles from free space A_i and from all other triangles j, S_{ij}. Whereas the impinging particles are incorporated into the growing layer with a sticking coefficient sc, the remaining particles are desorbed and may arrive at other parts of the surface. Here, the notation from the literature [132] is used, in which also more details are given:

$$R_i = \mathrm{sc} \cdot \left(A_i + \sum_j S_{ij} \right) \quad (35.105)$$

Here, the fluxes from free space and from all other triangles are calculated based on the solid angle for free sight to gas space Ω_{free}, the solid angle of free sight to all other triangles j, $\Delta\Omega_{ij}$, and the angle ϑ_{ij} between the surface normal of triangle i and the straight line between the centers of triangle i and triangle j; see Fig. 35.24:

$$S_{ij} = R_j \frac{1 - \mathrm{sc}}{\pi\ \mathrm{sc}} \cos\vartheta_{ij} \cdot \Delta\Omega_{ij} \quad (35.106)$$

The flux A_i from the gas volume is obtained by

$$A_i = G \int_{\Omega_{\mathrm{free}}} \cos\vartheta\, \mathrm{d}\Omega \quad (35.107)$$

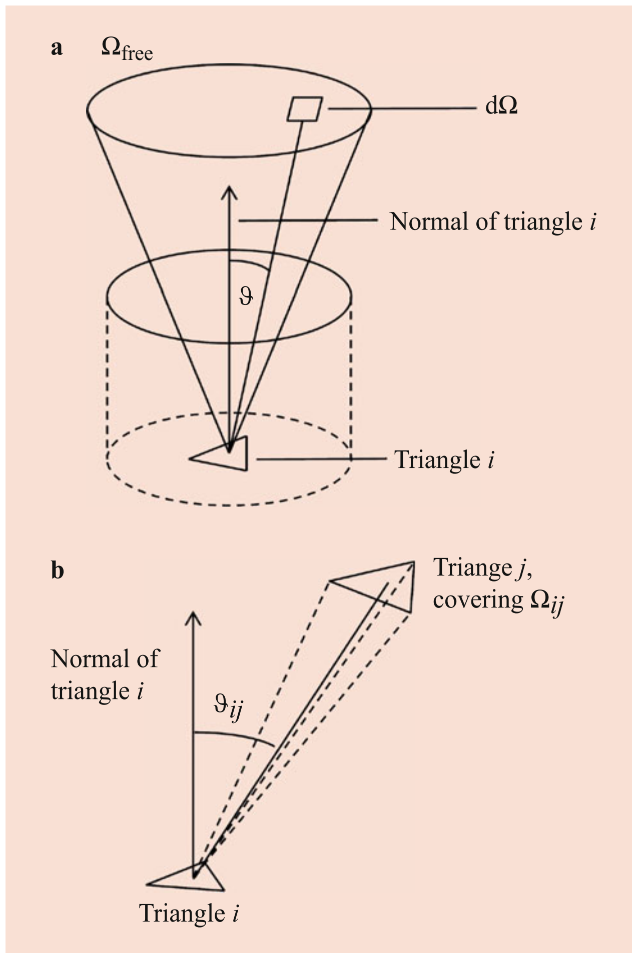


Fig. 35.24 (**a**) Solid angle Ω_{free} covering the region of free sight to gas space for a triangle i at the bottom of a cylindrical contact hole; (**b**) geometry for two arbitrary triangles with particle transfer to be calculated

where ϑ is the angle between the normal of triangle i and the direction from triangle i to dΩ (see Fig. 35.24).

Establishing the flux balance between all triangles and with free space results in a system of linear equations for the reaction fluxes R_i, with the transfer matrix T_{ij} between triangles I and j:

$$\pi R_i - \sum_{j \neq i} T_{ij} R_j = \pi - \frac{1}{1 - \text{sc}} \sum_{j \neq i} T_{ij} \tag{35.108}$$

$$T_{ij} = (1 - \text{sc}) \cos \vartheta_{ij} \cdot \Delta\Omega_{ij} \tag{35.109}$$

Here, for the transfer matrix elements T_{ij}, the calculation of the angles of free sight to gas space and especially to all other triangles is needed. For the latter, it must be checked if the view is (partly) blocked by any other triangle. In turn, for a surface discretization consisting of n triangles, the numerical effort scales with the third power of n.

Finally, the absolute deposition rates for each surface segment are obtained by normalizing to the deposition rate for a one-dimensional (flat) geometry. That one-dimensional deposition rate and the sticking coefficient are the only physical parameters of the process, whereas the transfer matrix characterizes the geometry. These have to be re-calculated after each time step to take the changes of the geometry into account.

In terms of computational efficiency, subsequent work has dealt especially with the optimum way to calculate the angles of free sight. Other numerical issues have been the appropriate adaptive surface discretization with one of the methods outlined in Sect. 35.6.2 and the efficient implementation of surface movement.

Figure 35.23 shows an example of a simulation that has been carried out with the model described above assuming a sticking coefficient of 0.15 which is a typical value for the modeling of a low-temperature oxide (LTO) process.

This basic model for deposition has been extended and adapted to the different deposition processes in various respects:

- Processes in which more than one mechanism contributes to the deposition. In the simplest case, the approach is applied simultaneously to two or more reaction species, which show no interaction and are each described by their one-dimensional deposition rate and their sticking coefficient.
- Long-throw sputter deposition which is characterized by the emission characteristics of the sputter source, collisionless movement of the sputtered particles due to low pressure, and a sticking coefficient of unity. Here, the transfer matrix is not needed, and the local deposition rates are calculated by integrating the contributions of the sputter source across the angle of free sight for the surface point in question, as discussed in the literature [134].
- Considering different species which either react on the surface and/or sputter away some already deposited atoms. This leads to additional terms in Eq. (35.105) while not changing the basic approach because the role and computation of the transfer matrix does not change. An example is ionized metal plasma deposition [135].

Such models can also be implemented by using Monte Carlo ray tracing instead of the transfer matrix method [136]: Here, trajectories of pseudo-particles are traced until they hit the layer surface. Similar to above, physical models are employed for the initial emission statistics of the particles and their adsorption, reaction, and desorption at the wafer surface. This also allows for the simulation of deposition

processes which are carried out in two steps, like atomic layer deposition (ALD) [137].

For etching, depending on the process in question, some or all of the elements discussed above for deposition may show up. In most cases of etching, physical and chemical sputtering play major roles, with the corresponding yields introduced as model parameters, replacing or complementing sticking coefficients [138]. However, adsorption/desorption and in turn transfer matrices may still play a role, because for some etching processes the emission and re-adsorption of species need to be modeled.

A detailed open-access review paper [139] deals especially with numerical aspects linked to the usage of the level set algorithm and physical models for etching simulation.

35.7 Process Variations

For advanced micro- and nanoelectronic devices and circuits, both statistical and systematical process variations come into play, which influence the performance of devices and systems. In result, although the nominal product, fabricated under ideal circumstances, may meet the specifications and required figures of merit, a considerable fraction of devices and circuits may fail to do so, causing yield loss and critical increase of fabrication costs per sold unit. In turn, the impact of such process variations must be known and minimized as much as necessary.

Statistical process variations such as random dopant fluctuations RDF [3], line edge roughness LER [140], and metal grain granularity MGG [141] result from the granularity of matter. They have been considered since long in dedicated device studies, employing special efficient device simulation tools [142] and simple but valid assumptions on the statistical distribution of these variations. This is in line with the frequent practice in device simulation to start from assumed idealized device geometries and dopant distributions. However, in a real fabrication environment not only the full process flow must be considered, which leads to non-idealized device geometries and dopant distributions, but also its variations. In the following, a brief overview of the sources of systematic process variations is given, together with a hierarchical approach to simulate their effect on devices and circuits.

35.7.1 Sources of Systematic Process Variations

Systematic process variations may occur in all major process steps. They can be grouped into two categories: First, inhomogeneity which is inherent to the equipment used and which can neither be completely avoided by optimizing the equipment design nor the process flow. Second, some process parameters cannot be controlled precisely or may have some drifts during processing or on the long run between repeated process steps.

For both optical and EUV lithography, both the distance between the optical system (last lens or last mirror) and the wafer and the energy emitted by the light source may change from illumination to illumination in the standard step-and-repeat process, leading to systematic variations between different dies on one wafer and especially between different wafers. The so-called process window in Fig. 35.25 shows examples how these variations of focus and dose, respectively, modify the size of the features generated in the lithography step, the so-called critical dimensions (CD).

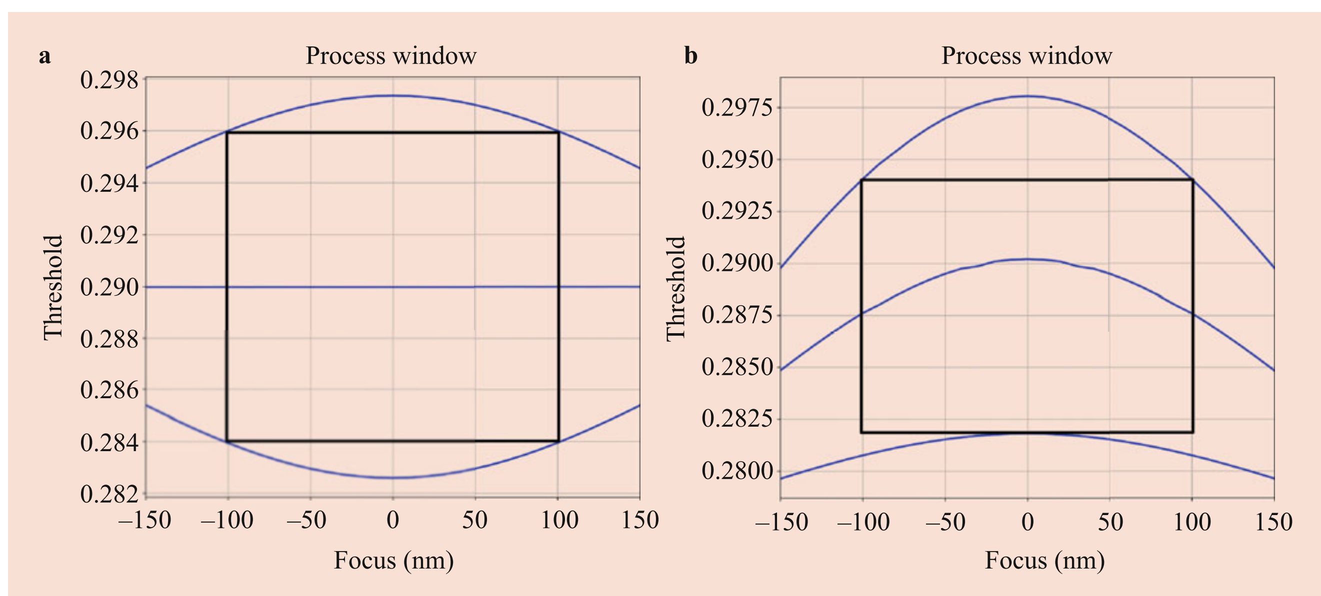


Fig. 35.25 Process window in optical lithography for (**a**) 120 nm pitch with 60 nm lines and (**b**) 72 nm lines

Here, the central line shows the combinations of focus and dose which lead to the nominal CD (60 nm left picture, 72 nm right picture), whereas the upper and lower lines show the combinations of focus and dose which lead to a CD increase or decrease by 10%, respectively. Whereas under specific imaging conditions, the CD stays constant for varying focus at a specific fixed dose, for other imaging conditions, this is not the case. Additionally, also some other process variations may occur in lithography steps, caused, e.g., by defects and imperfections of the optical system or imperfect wafer alignment.

For etching and deposition, especially inhomogeneous concentrations of ions and neutrals, the temperature distribution, or the emission characteristics of the sputter target lead to systematic process variations across the wafers or between wafers. Figure 35.26 shows an example for the dependence of the etch rate on the distance from the center of the wafer for reactive ion etching of silicon based on chlorine chemistry.

Further systematic process variations include among others differences between nominally equal temperature profiles in very short time annealing processes, such as millisecond or laser thermal annealing [130].

35.7.2 Hierarchical Simulation of the Impact of Process Variations

In order to predict the impact of process variations on devices and systems, it is necessary to trace them through the whole fabrications process: Not only the process step where the variation occurs must be simulated, but all subsequent process steps must be simulated, using as starting conditions both the intermediate results without and with the effects of that variation. Furthermore, it is necessary to make sure that numerical errors which occur throughout the simulation or when changing between the data representation of different simulation modules or tools do not invalidate the variability study: They must affect the result significantly less than the variation in question.

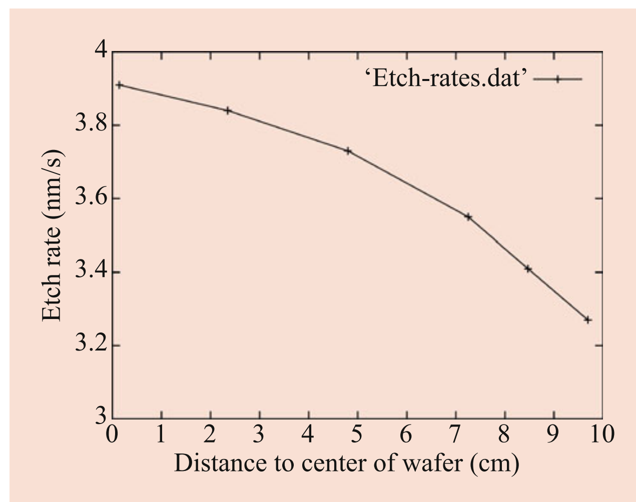


Fig. 35.26 Example for the dependence of etch rate of a silicon etching process on the distance from the center of the wafer

From the modeling point of view, the study of process variations requires to employ appropriate models: First, models which describe the effect of the process variation at the step at which it occurs – e.g., the impact of focus on the CD generated in a lithography step. Next, models must be used which are capable of tracing the changes of the results of that process step (here the CD) through all following process steps. Generally, both requirements can be met by selecting a suitable process model among the ones available.

In real cases, a state-of-the-art process flow may be affected by a large number of process variations. Considering just 10 process variations and just 5 parameter values for each of them would yield a final split of 5^{10} or nearly ten million process simulations, which is of course not feasible. In turn, it is necessary to first select the most relevant process variations (about 3–5) and then to employ a suitable design-of-experiment approach to define the set of process simulations needed, as discussed elsewhere [143].

In result, geometries, dopant distributions, and potentially other quantities which characterize a device (or interconnect) are simulated for different values of the input process variations. Considering, e.g., three process variations which are discretized with three values each, a matrix of $3 \times 3 \times 3$ devices results. In order to extract compact models which are aware of variations, first the nominal compact model for the nominal device without any variations is extracted. In the second step, this compact model is extended to also include the process corners considered, which means in the example addressed here the $3 \times 3 \times 3$ devices [144]. Finally, statistical device simulation is employed to extend this compact model to also include the statistical process variations, such as RDF [144]. A later improvement of this method [145] has enabled the direct usage of the values of the systematic process variations (e.g., defocus) in the second compact model extraction step outlined above [143]; see Fig. 35.27. As an example Fig. 35.28 shows the saturation current of a nanowire transistor.

More detailed descriptions of the sources of process variations and of the hierarchical simulation of their impacts on devices and are given in a related open-access papers [143] [145].

35.8 Conclusions

Process simulation is the virtual image of IC manufacturing on a computer: It covers the whole development from the bare silicon wafer to the final device and circuit. Its full benefit can only materialize if accurate physical models with proper parameters are available within numerically stable and correct simulators.

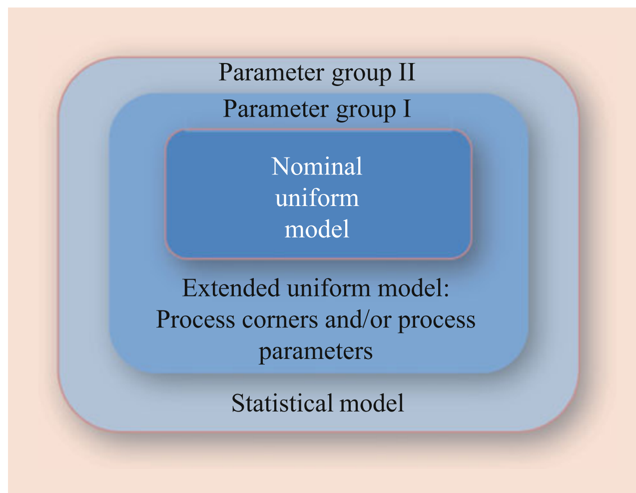


Fig. 35.27 Extraction and generation of hierarchical compact model aware of systematic and statistical process variations, following the approach published before [145]

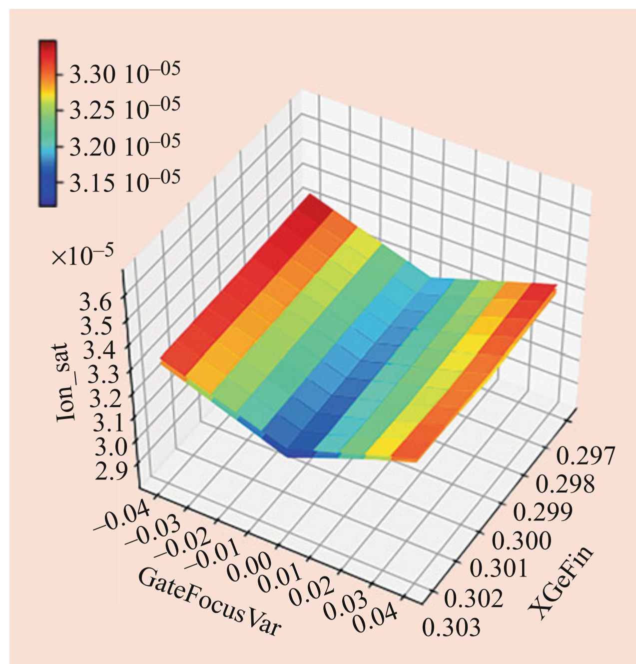


Fig. 35.28 Saturation current (color in amperes) of a nanowire NMOS described by the extended compact model (color in amperes) [145]

The applications of process simulation are manifold. Nowadays it is most frequently used in industry for the development and optimization of the process flows needed for the fabrication of devices. The success of this approach depends on the availability of advanced process simulation tools which cover the whole fabrication sequence, are closely linked with device simulation, and are accompanied by good user support. Because the speed of simulation is a key requirement for the simulation splits used in technology and device optimization, very frequently efficient models are used which depend on calibrations for the process flow in question. On the other hand, physically rigorous models are readily used for development and optimization of new advanced processes or pieces of equipment. Here, predictivity with as little calibration as possible is the key requirement, not the integration into an overall simulation system.

An increasingly important advantage of process simulation is that it can be used not only to study and select process options but also to quantify and minimize the impact of systematic process variations which are caused by the equipment used. Here, it makes indispensable contributions to the simultaneous development and optimization of technologies, devices, circuits, and designs, the so-called Design-Technology Co-Optimization (DTCO), which has in recent year developed into a key and indispensable method for the nanoelectronics industry.

With new semiconductor materials coming into use and new device architectures (including new carriers of information like spin or the phase of the material) emerging, the complexity of the processes to be simulated is continuing to grow. So are the requirements on model accuracy and generality and efficiency of simulation. There will also in future be several (but of course different from the past) challenges left for developers in the process simulation area.

Acknowledgments The authors of this chapter want to acknowledge the funding of some of the research leading to their results presented here by the European Union and by the German Federal Ministry of Education and Research. They want to acknowledge the contributions of colleagues from Fraunhofer IISB in terms of discussions and preparation of some figures, especially by E. Bär, A. Burenkov, and P. Evanschitzky.

References

1. 2011 International Technology Roadmap for Semiconductors, Modeling and Simulation, Table MS3, http://www.itrs2.net/2011-itrs.html
2. 2013 International Technology Roadmap for Semiconductors, Lithography, https://www.semiconductors.org/resources/2013-international-technology-roadmap-for-semiconductors-itrs/
3. Asenov, A.: Simulation of statistical variability in nano MOSFETs. In: Proceedings IEEE Symposium VLSI Technology, June 12–14, 2007, pp. 86–87
4. Selberherr, S.: Analysis and Simulation of Semiconductor Devices. Springer, Wien, New York (1984)
5. Pichler, P.: Intrinsic Point Defects, Impurities, and their Diffusion in Silicon. Springer, Wien-New York (2004)
6. D.A. Antoniadis, S.E. Hansen, R.W. Dutton, A.G. Gonzalez: SUPREM 1 – A program for IC processing and Simulation, Stanford University Technical report No. 5019-1, Stanford (1977)
7. Antoniadis, D.A., Dutton, R.W.: Models for computer simulation of complete IC fabrication process. Trans. Electron Devices. **ED-26**(4), 490–500 (1979)
8. Ryssel, H., Haberger, K., Hoffmann, K., Prinke, G., Dümcke, R., Sachs, A.: Simulation of doping processes. IEEE Trans. Electron Devices. **ED-27**(8), 1484–1492 (1980)

9. M.E Law, C.S. Rafferty, R.W. Dutton, New N-well fabrication techniques based on 2D process simulation, Proc. IEDM 1986, 518-521 (1986)
10. Lorenz, J., Pelka, J., Ryssel, H., Sachs, A., Seidl, A., Svoboda, M.: COMPOSITE – a complete modeling program of silicon technology. IEEE Trans. Electron Devices. **32**(10) (1977-1986)
11. Sentaurus Process, www.synopsys.com
12. Victory Process, www.silvaco.com
13. Bottinger, J., Davies, J.A., Siegmund, P., Winterbon, K.B.: On reflection coefficient of keV heavy-ions beams from solid targets. Rad. Effects. **11**(2), 69 (1971)
14. Lorenz, J., Krüger, W., Barthel, A.: Simulation of the lateral spread of implanted ions: theory. In: Miller, J.J.H. (ed.) Proceedings Sixth International NASECODE Conference, pp. 513–520. Boole Press, Dublin (1989)
15. Hofker, W.K.: Concentration profiles of boron implantations in amorphous and polycrystalline silicon. Philips Res. Rep. **8**, 41 (1975)
16. Ryssel, H., Ruge, I.: Ion Implantation. Wiley, Chichester/New York/Brisbane/Toronto/Singapore (1986)
17. Tasch, A.F., Shin, H., Park, C.: An improved approach to accurately model Shallow B and BF_2 implants in silicon. J. Electrochem. Soc. **136**(3), 810 (1989)
18. Ishiwara, H., Furakawa, S., Yamada, J., Kawamura, M.: S. Namba, Ion Implantation in Semiconductors, p. 423. Plenum Press, New York (1975)
19. Ryssel, H., Krüger, W., Lorenz, J.: Comparison of Monte-Carlo simulations and analytical models for the calculation of ion implantation profiles in multilayer targets. Nucl. Instrum. Methods Phy. Res. **B19**(20), 40–44 (1987)
20. Wiezbicki, R.J., Lorenz, J., Barthel, A.: Simulation of ion implantation into multilayer structures. In: Heuberger, A., Ryssel, H., Lange, P. (eds.) Proceedings ESSDERC 1989, pp. 193–197. Springer, Berlin, Heidelberg (1989)
21. Webb, R.P., Maydell, E.: Comparison of fast algorithms for calculation of range profiles in layered structures. Nucl. Instr. Methods. **B33**, 117–121 (1988)
22. Wierzbicki, R.J.: Analytische Beschreibung der Implantation von Ionen in Ein- und Mehrschichtstrukturen. Shaker Verlag, Aachen (1994)
23. Wiezbicki, R.J., Biersack, J.P., Barthel, A., Lorenz, J., Ryssel, H.: Reflection approach for the analytical description of light ion implantation into bilayer structure. Radiat. Eff. Solids. **130–131**, 495–506 (1994)
24. Furukawa, S., Matsumura, M., Ishiwara, H.: Theoretical considerations on lateral spread of implanted ions. Jap. J. Appl. Phys. **11**(2), 134 (1972)
25. Ryssel, H., Gong, L., Lorenz, J.: Improvements in simulation of implantation profiles. In: Proceedings 1989 International Symposium on VLSI Technology, Systems and Applications, pp. 102–105, Taipeh, 17–19 May 1989
26. Lorenz, J., Barthel, A., Gong, L., Ryssel, H., Wierzbicki, R.J.: Analytical description of ion implantation profiles. In: Barraclough, K., Chikawa, J., Huff, H. (eds.) Proceedings 6th International Symposium on Silicon Materials Science and Technology, pp. 538–549. Electrochemical Society, Pennington (1989)
27. Lorenz, J., Wiezbicki, R.J., Ryssel, H.: Analytical modeling of lateral implantation profiles. Nucl. Inst. Methods Phys. Res. B. **96**, 168–172 (1995)
28. Cole, P.D., Crean, G.M., Lorenz, J., Dupas, L.: Comparison of models for the calculation of ion implantation moments of implanted boron, phosporus and arsenic dopants in thin film silicides. Nucl. Inst. Methods Phys. Res. B. **55**, 763–768 (1991)
29. Biersack, J.P., Haggmark, L.G.: A Monte-Carlo computer program for the transport of energetic ions in amorphous targets. Nucl. Instrum. Methods Phy. Res. **174**, 257–269 (1980)
30. Hobler, G., Selberherr, S.: Two-dimensional modeling of ion implantation induced point defects. IEEE Trans. Comput.-Aid Design. **7**(3), 174–180 (1988)
31. Ziegler, J.: Ion implantation physics. In: Ziegler, J.F. (ed.) Ion Implantation Science and Technology, pp. 175–238. Ion Implantation Technology Co., Edgewater (2000)
32. Robinson, M.T., Torrens, I.M.: Computer simulation of atomic-displacement cascades in solids in the binary-collision approximation. Phys. Rev. **B9**(12), 5008–5023 (1974)
33. Norgett, M.J., Robinson, M.T., Torrens, I.M.: A proposed method of calculating displacement dose rates. Nucl. Eng. Des. **33**, 50–54 (1975)
34. Pichler, P., Sledziewski, T., Häublein, V., Bauer, A.J., Erlbacher, T.: Channeling in 4H-SiC from an application point of view. Mater. Sci. Forum. **963**, 386–389 (2019)
35. Burenkov, A., Tietzel, K., Hössinger, A., Lorenz, J., Ryssel, H., Selberherr, S.: A computationally efficient method for three-dimensional simulation of ion implantation. IEICE Trans. Electron. **E83-C**(8), 1259–1265 (2000)
36. Bohmayr, W., Burenkov, A., Lorenz, J., Ryssel, H., Selberherr, S.: Trajectory split method for Monte Carlo simulation of ion implantation. IEEE Trans. Semicond. Manuf. **8**(4), 402–407 (1995)
37. Tian, X.B., Kwok, D.T.W., Chun, P.K.: Modeling of incident particle energy distribution in plasma immersion ion implantation. J. Appl. Phys. **88**(9), 4961–4966 (2000)
38. Burenkov, A., Hahn, A., Spiegel, Y., Etienne, H., Torregrosa, F.: Simulation of BF_3 plasma immersion ion implantation into silicon. In: Pelaz, L., Duffy, R., Torregrosa, F., Bourdelle, K. (eds.) Ion Implantation Technology 2012, pp. 233–236. AIP Conference Proceedings, Melville (2012)
39. Fahey, P.M., Griffin, P.B., Plummer, J.D.: Point defects and dopant diffusion in silicon. Rev. Mod. Phys. **61**(2), 289–384 (1989)
40. Pelaz, L., Marqués, L.A., Aboy, M., López, P., Santos, I.: Front-end process modeling in silicon. Eur. Phys. J. B. **72**(3), 323–359 (2009)
41. Bracht, H.: Self-and dopant diffusion in silicon, germanium, and their alloys. In: Kissinger, I.G., Pizzini, S. (eds.) Silicon, Germanium, and Their Alloys: Growth, Defects, Impurities, and Nanocrystals, pp. 159–206. CRC Press (2014)
42. Voronkov, V., Falster, R.: Multiple structural forms of a vacancy in silicon as evidenced by vacancy profiles, produced by rapid thermal annealing. Phys. Status Solidi B. **251**(11), 2179–2184 (2014)
43. Mikkelsen Jr., J.C.: The diffusivity and solubility of oxygen in silicon. Mat. Res. Soc. Symp. Proc. **59**, 19–30 (1986)
44. Windl, W., Bunea, M.M., Stumpf, R., Dunham, S.T., Masquelier, M.P.: First-principles study of boron diffusion in silicon. Phys. Rev. Lett. **83**(21), 4345 (1999)
45. De Salvador, D., Napolitani, E., Mirabella, S., Bisognin, G., Impellizzeri, G., Carnera, A., Priolo, F.: Atomistic mechanism of boron diffusion in silicon. Phys. Rev. Lett. **97**, 255902 (2006)
46. Yoshida, M., Arai, E., Nakamura, H., Terunuma, Y.: Excess vacancy generation mechanisms at phosphorus diffusion into silicon. J. Appl. Phys. **45**(4), 1498–1506 (1974)
47. Seeger, A., Chik, K.P.: Diffusion mechanism and point defects in silicon and germanium. Phys. Status Solidi. **29**, 455–542 (1968)
48. Gösele, U., Frank, W., Seeger, A.: Mechanism and kinetics of the diffusion of gold in silicon. Appl. Phys. **23**, 361–368 (1980)
49. Frank, F.C., Turnbull, D.: Mechanism of diffusion of copper in germanium. Phys. Rev. **104**(3), 617–618 (1956)
50. Waite, T.R.: Theoretical treatment of the kinetics of diffusion-limited reactions. Phys. Rev. **107**(2), 463–470 (1957)
51. Debye, P.: Reaction rates in ionic solutions. Trans. Electrochem. Soc. **82**, 265–272 (1942)
52. Waite, T.R.: General theory of bimolecular reaction rates in solids and liquids. J. Chem. Phys. **28**(1), 103–106 (1958)

53. Cowern, N.E.B., Janssen, K.T.F., van de Walle, G.F.A., Gravesteijn, D.J.: Impurity diffusion via an intermediate species: the B-Si system. Phys. Rev. Lett. **65**(19), 2434–2437 (1990)
54. Cowern, N.E.B.: General model for intrinsic dopant diffusion in silicon under nonequilibrium point-defect conditions. J. Appl. Phys. **64**(9), 4484–4490 (1988)
55. Uppal, S., Willoughby, A.F.W., Bonar, J.M., Cowern, N.E.B., Grasby, T., Morris, R.J.H., Dowsett, M.G.: Diffusion of boron in germanium at 800–900°C. J. Appl. Phys. **96**(3), 1376–1380 (2004)
56. Orr Arienzo, W.A., Glang, R., Lever, R.F., Lewis, R.K., Morehead, F.F.: Boron diffusion in silicon at high concentrations. J. Appl. Phys. **63**(1), 116–120 (1988)
57. Smits, F.M.: Formation of junction structures by solid-state diffusion. Proc. IRE. **46**, 1049–1061 (1958)
58. Fair, R.B., Tsai, J.C.C.: A quantitative model for the diffusion of phosphorus in silicon and the emitter dip effect. J. Electrochem. Soc. **124**(7), 1107–1117 (1977)
59. Nylandsted Larsen, A., Kyllesbech Larsen, K., Andersen, P.E., Svensson, B.G.: Heavy doping effects in the diffusion of group IV and V impurities in silicon. J. Appl. Phys. **73**(2), 691–698 (1993)
60. Mathiot, D., Pfister, J.C.: Diffusion of arsenic in silicon: validity of the percolation model. Appl. Phys. Lett. **42**(12), 1043–1044 (1983)
61. Ramamoorthy, M., Pantelides, S.T.: Complex dynamical phenomena in heavily arsenic doped silicon. Phys. Rev. Lett. **76**(25), 4753–4756 (1996)
62. Wills, G.N.: The orientation dependent diffusion of boron in silicon under oxidizing conditions. Solid State Electron. **12**, 133–134 (1969)
63. Antoniadis, D.A., Gonzales, A.G., Dutton, R.W.: Boron in near-intrinsic <100> and <111> silicon under inert and oxidizing ambients – diffusion and segregation. J. Electrochem. Soc. **125**(5), 813–819 (1978)
64. Masetti, G., Solmi, S., Soncini, G.: On phosphorus diffusion in silicon under oxidizing atmospheres. Solid State Electron. **16**, 1419–1421 (1973)
65. Mizuo, S., Higuchi, H.: Retardation of Sb diffusion in Si during thermal oxidation. Jpn. J. Appl. Phys. **20**(4), 739–744 (1981)
66. Dobson, P.S.: The effect of oxidation on anomalous diffusion in silicon. Philos. Mag. **24**, 567–576 (1971)
67. Hu, S.M.: Formation of stacking faults and enhanced diffusion in the oxidation of silicon. J. Appl. Phys. **45**(4), 1567–1573 (1974)
68. Taniguchi, K., Shibata, Y., Hamaguchi, C.: Theoretical model for self-interstitial generation at the Si/SiO_2 interface during thermal oxidation of silicon. J. Appl. Phys. **65**(7), 2723–2727 (1989)
69. Mizuo, S., Kusaka, T., Shintani, A., Nanba, M., Higuchi, H.: Effect of Si and SiO_2 thermal nitridation on impurity diffusion and oxidation induced stacking fault size in Si. J. Appl. Phys. **54**(7), 3860–3866 (1983)
70. Ahn, S.T., Kennel, H.W., Plummer, J.D., Tiller, W.A.: Film stress-related vacancy supersaturation in silicon under low-pressure chemical vapor deposited silicon nitride films. J. Appl. Phys. **64**(10), 4914–4919 (1988)
71. Cowern, N.E.B.: Diffusion in a single crystal within a stressed environment. Phys. Rev. Lett. **99**, 155903 (2007)
72. Sentaurus Process of Synopsys, Inc., with Advanced Calibration. Version P-2019.03, 2019
73. Aboy, M., Santos, I., Pelaz, L., Marqués, L.A., López, P.: Modeling of defects, dopant diffusion and clustering in silicon. J. Comput. Electron. **13**, 40–58 (2014)
74. Napolitani, E., Impellizzeri, G.: Ion implantation defects and shallow junctions in Si and Ge. In: Romano, L., Privitera, V., Jagadish, C. (eds.) Defects in Semiconductors, pp. 93–122. Academic Press (2015)
75. Kim, Y., Massoud, H.Z., Gösele, U.M., Fair, R.B.: Physical modeling of the time constant of transient enhancement in the diffusion of ion-implanted dopants in silicon. Electrochem. Soc. Proc. **91-4**, 254–272 (1991)
76. Eaglesham, D.J., Stolk, P.A., Gossmann, H.-J., Poate, J.M.: Implantation and transient B diffusion in Si: the source of the interstitials. Appl. Phys. Lett. **65**(18), 2305–2307 (1994)
77. Bonafos, C., Omri, M., de Mauduit, B., BenAssayag, G., Claverie, A., Alquier, D., Martinez, A., Mathiot, D.: Transient enhanced diffusion of boron in presence of end-of-range defects. J. Appl. Phys. **82**(6), 2855–2861 (1997)
78. Cowern, N.E.B., Mannino, G., Stolk, P.A., Roozeboom, F., Huizing, H.G.A., van Berkum, J.G.M., Cristiano, F., Claverie, A., Jaraíz, M.: Energetics of self-interstitial clusters in Si. Phys. Rev. Lett. **82**(22), 4460–4463 (1999)
79. Claverie, A., Colobeau, B., De Mauduit, B., Bonafos, C., Hebras, X., Assayag, G.B., Cristiano, F.: Extended defects in shallow implants. Appl. Phys. A Mater. Sci. Process. **76**, 1025–1033 (2003)
80. Zechner, C., Zographos, N., Matveev, D., Erlebach, A.: Accurate and efficient TCAD model for the formation and dissolution of small interstitial clusters and {311} defects in silicon. Mater. Sci. Eng. B. **401-403**, 124–125 (2005)
81. Zographos, N., Zechner, C., Avci, I.: Efficient TCAD model for the evolution of interstitial clusters, {311} defects, and dislocation loops in silicon. Mat. Res. Soc. Symp. Proc. **994**, 0994-F10-01 (2007)
82. Wolf, F.A., Martinez-Limia, A., Stichtenoth, D., Pichler, P.: Modeling the annealing of dislocation loops in implanted c-Si solar cells. IEEE J. Photovoltaics. **4**(3), 851–858 (2014)
83. Giles, M.: Transient phosphorus diffusion below the amorphization threshold. J. Electrochem. Soc. **139**(4), 1160–1165 (1991)
84. Pelaz, L., Gilmer, G.H., Jaraiz, M., Herner, S.B., Gossmann, H.-J., Eaglesham, D.J., Hobler, G., Rafferty, C.S., Barbolla, J.: Modeling of the ion mass effect on transient enhanced diffusion: deviation from the "+1" model. Appl. Phys. Lett. **73**(10), 1421–1423 (1998)
85. Hobler, G., Moroz, V.: Initial conditions for transient enhanced diffusion: beyond the plus-factor approach. In: Tsoukalas, D., Tsamis, C. (eds.) Simulation of Semiconductor Processes and Devices 2001, pp. 34–37. Springer-Verlag (2001)
86. Schwenker, R.O., Pan, E.S., Lever, R.F.: Arsenic clustering in silicon. J. Appl. Phys. **42**(8), 3195–3200 (1971)
87. Hu, S.M.: Diffusion in silicon and germanium. In: Shaw, D. (ed.) Atomic Diffusion in Semiconductors, pp. 217–350. Plenum Press (1973)
88. Pelaz, L., Gilmer, G.H., Gossmann, H.-J., Rafferty, C.S., Jaraiz, M., Barbolla, J.: B cluster formation and dissolution in Si: a scenario based on atomistic modeling. Appl. Phys. Lett. **74**(24), 3657–3659 (1999)
89. Harrison, S.A., Edgar, T.F., Hwang, G.S.: Interstitial-mediated arsenic clustering in ultrashallow junction formation. Electrochem. Solid-State Lett. **9**(12), G354–G357 (2006)
90. Massalski, T.B., Okamoto, H., Subramanian, P.R., Kacprzak, L. (eds.): Binary Alloy Phase Diagrams, 2nd edn. ASM International, Materials Park (1990)
91. Nobili, D., Solmi, S., Parisini, A., Derdour, M., Armigliato, A., Moro, L.: Precipitation, aggregation, and diffusion in heavily arsenic-doped silicon. Phys. Rev. B. **49**(4), 2477–2483 (1994)
92. Bourret, A., Schröter, W.: HREM of SiP precipitates at the (111) silicon surface during phosphorus predeposition. Ultramicroscopy. **14**, 97–106 (1984)
93. Dabrowski, J., Müssig, H.-J., Zavodinsky, V., Baierle, R., Caldas, M.J.: Mechanism of dopant segregation to SiO_2/Si(001) interfaces. Phys. Rev. B. **65**, 245305 (2002)
94. Steen, C., Martinez-Limia, A., Pichler, P., Ryssel, H., Paul, S., Lerch, W., Pei, L., Duscher, G., Severac, F., Cristiano, F., Windl,

W.: Distribution and segregation of arsenic at the SiO_2/Si interface. J. Appl. Phys. **104**, 023518 (2008)
95. Pei, L., Duscher, G., Steen, C., Pichler, P., Ryssel, H., Napolitani, E., De Salvador, D., Piro, A.M., Terrasi, A., Severac, F., Cristiano, F., Ravichandran, K., Gupta, N., Windl, W.: Detailed arsenic concentration profiles at Si/SiO_2 interfaces. J. Appl. Phys. **104**, 043507 (2008)
96. Duffy, R., Venezia, V.C., Heringa, A., Hüsken, T.W.T., Hopstaken, M.J.P., Cowern, N.E.B., Griffin, P.B., Wang, C.C.: Boron uphill diffusion during ultrashallow junction formation. Appl. Phys. Lett. **82**(21), 3647–3649 (2003)
97. Lau, F., Mader, L., Mazure, C., Werner, C., Orlowski, M.: A model for phosphorus segregation at the silicon-silicon dioxide interface. Appl. Phys. A Mater. Sci. Process. **49**, 671–675 (1989)
98. Orlowski, M.: New model for dopant redistribution at interfaces. Appl. Phys. Lett. **55**(17), 1762–1764 (1989)
99. Jaraiz, M., Pelaz, L., Rubio, E., Barbolla, J., Gilmer, G.H., Eaglesham, D.J., Gossmann, H.J., Poate, J.M.: Atomistic modeling of point and extended defects in crystalline materials. Mat. Res. Soc. Symp. Proc. **532**, 43–53 (1998)
100. Martin-Bragado, I., Moroz, V.: Facet formation during solid phase epitaxy regrowth: a lattice kinetic Monte Carlo model. Appl. Phys. Lett. **95**, 123123 (2009)
101. Deal, B.E., Grove, A.S.: General relationship for the thermal oxidation of silicon. J. Appl. Phys. **36**(12), 3770–3778 (1965)
102. Massoud, H.Z., Plummer, J.D.: Thermal oxidation of silicon in dry oxygen growth-rate enhancement in the thin regime. J. Electrochem. Soc. **132**(11), 2685–2693 (1985)
103. Kao, D.-B., McVittie, J.P., Nix, W.D., Saraswat, K.C.: Two-dimensional silicon oxidation experiments and theory. In: Proceedings IEDM 1985, pp. 388–391. IEEE (1985)
104. Seidl, A.: Zweidimensionale Simulation der lokalen Oxidation von Silicium, PhD thesis, Universität Erlangen-Nürnberg, 1988
105. Eyring, H.: Viscosity, plasticity and diffusion as examples of absolute reaction rates. J. Chem. Phys. **4**, 283–291 (1936)
106. Rafferty, C.S., Borucki, L., Dutton, R.W.: Plastic flow during thermal oxidation of silicon. Appl. Phys. Lett. **54**, 1516–1518 (1989)
107. Chin, D., Oh, S.-Y., Hu, S.-M., Dutton, R., Moll, J.L.: Two-dimensional oxidation. IEEE Trans. Electron Devices. **ED-30 (7)**, 744–749 (1983)
108. Chorin, A.J.: A numerical method for solving incompressible viscous flow problems. J. Comput. Phys. **2**, 12–26 (1967)
109. Mack, C.: Fundamental Principles of Optical Lithography. WILEY, The Atrium/Southern Gate/Chichester/West Sussex/England (2007)
110. Multiple Patterning, Semiconductor Engineering, https://semiengineering.com (2019)
111. Wong, A.K.K.: Optical Imaging in Projection Microlithography. SPIE Press, Bellingham (2005)
112. Bakshi, V.: EUV Lithography. SPIE Press, Bellingham (2017)
113. Hopkins, H.H.: On the diffraction theory of optical image. Proc. R. Soc. London Ser. A. **217**, 408–432 (1953)
114. Born, M., Wolf, E.: Principles of Optics: Electromagnetic Theory of Propagation, Interference, and Diffraction of Light. Cambridge University Press, London (1999)
115. Taflove, A.: The Finite-Difference Time-Domain Method. ARTECH HOUSE, Boston (1995)
116. Moharam, M.G., Pommet, D.A., Grann, E.B., Gaylord, T.K.: Stable implementation of the rigorous coupled wave analysis for surface relief gratings: enhanced transmittance matrix approach. J. Opt. Soc. Am. **A 12**, 1077–1086 (1995)
117. Erdmann, A., Evanschitzky, P., Citarella, G., Fühner, T.; De Bisschop, P.: Rigorous mask modeling using waveguide and FDTD methods: an assessment for typical hyper-NA imaging problems. In: Proceedings SPIE 6283, Photomask and Next-Generation Lithography Mask Technology XIII, Yokohama, 2006, 628319 (19 May 2006); https://doi.org/10.1117/12.681872
118. Dill, F.H., Hornberger, W.P., Hauge, P.S., Shaw, J.M.: Characterization of positive photoresist. IEEE Trans. Electron Devices. **ED-22**(7), 445–445 (1975)
119. PROLITH, www.kla-tencor.com
120. Sentaurus Litho, www.synopsys.com
121. OPTOLITH, www.silvaco.com
122. Fühner, T., Schnattinger, T., Ardelean, G., Erdmann, A.: Dr. LiTHO: a development and research lithography simulator. In: Proceedings SPIE 6520, Optical Microlithography XX, 65203F (27 March 2007); https://doi.org/10.1117/12.709535
123. Evanschitzky, P., Erdmann, A.: Advanced EUV mask and imaging modeling. J. Micro/Nanolith. MEMS MOEMS. **16**(4), 041005 (2017). https://doi.org/10.1117/1.JMM.16.4.041005
124. Fühner, T., Erdmann, A., Evanschitzky, P.: Simulation-based EUV source and mask optimization. In: Proceedings SPIE 7122, Photomask Technology 2008, 71221Y (17 October 2008); https://doi.org/10.1117/12.801436
125. Evanschitzky, P., Erdmann, A.: Efficient simulation of EUV pellicles. In: Proceedings SPIE 10450, International Conference on Extreme Ultraviolet Lithography 2017, 104500B (16 October 2017); https://doi.org/10.1117/12.2280535
126. Evanschitzky, P., Shao, F., Erdmann, A.: Efficient simulation of extreme ultraviolet multilayer defects with rigorous data base approach. J. Micro/Nanolith. MEMS MOEMS. **12**(2), 021005 (2013)
127. Maury, M., Hassouni, K., Michau, A., Torregrosa, F., Borvon, G.: Simulation of a PIII reactor with a magnetized remote source. In: Proceedings 20th International Symposium on Plasma Chemistry (ISPC 20), 2011
128. Simulator CFD-ACE+; ESI Group: Paris, France, 2018
129. Q-VT Plasma Processing Simulator; Quantemol Ltd.: London, UK, 2018
130. Hackenberg, M., Rommel, M., Rummler, M., Lorenz, J., Pichler, P., Huet, K., Negru, R., Fisicaro, G., la Magna, A., Taleb, N., Quillec, M.: Melt depth and time variations during pulsed laser thermal annealing with one and more pulses. In: Proceedings 43rd European Solid State Device Research Conference (ESSDERC 2013), pp. 214–217, Bucharest, Romania, 16–20 September, 2013, IEEE, Piscataway
131. Dill, F.H., Neureuther, A.R., Tuttle, J.A., Walker, E.J.: Modeling projection printing of positive photoresists. IBM Res. **RC 5261** (1975)
132. Bär, E., Lorenz, J.: 3-D simulation of LPCVD using segment-based topography discretization. IEEE Trans. Semicond. Manuf. **9**(1), 67–73 (1996)
133. Jewett, R.: A String Model Etching Algorithm, Memorandum N. UCB/ERL M79/68. Electronics Research Laboratory, University of California, Berkeley (1979)
134. Bär, E., Lorenz, J., Ryssel, H.: Simulation of the influence of via sidewall tapering on step coverage of sputter-deposited barrier layers. Microelectron. Eng. **64**, 321–328 (2002)
135. Kistler, S., Bär, E., Lorenz, J., Ryssel, H.: Three-dimensional simulation of ionized metal plasma vapor deposition. Microelectron. Eng. **76**, 100–105 (2004)
136. Rey, J.C., Cheng, L.-Y., McVittie, J.P., Saraswat, K.C.: Monte Carlo low pressure deposition profile simulations. J. Vac. Sci. Technol. A. **9**, 1083–1087 (1991)
137. Filipovic, L.: Modeling and simulation of atomic layer deposition. In: Proceedings SISPAD 2019, pp. 323–326. IEEE, Piscataway, 2019
138. Bär, E., Kunder, D., Evanschitzky, P., Lorenz, J.: Coupling of equipment simulation and feature-scale profile simulation for dry-etching of polysilicon gate lines. In: Baccarani, G., Rudan, M.

(eds.) Proceedings SISPAD 2010, pp. 57–60. IEEE, Piscataway (2010)
139. Klemenschits, X., Selberherr, S., Filipovic, L.: Modeling of gate stack patterning for advanced technology nodes: a review. Micromachines. **9**(12), 631 (2018). https://doi.org/10.3390/mi9120631
140. Asenov, A., Kaya, S., Brown, A.R.: Intrinsic parameter fluctuations in dacananometer MOSFETs introduced by gate line edge roughness. IEEE Trans. Electron Devices. **50**(5), 1254–1260 (2003)
141. Brown, A.R., Idris, N.M., Watling, J.R., Asenov, A.: Impact of metal grain granularity on threshold voltage variability: a full scale statistical simulation study. IEEE Electron Device Lett. **31**(11), 1199–1202 (2010)
142. Garand User Guide, {online}, https://solvnet.synopsys.com, Synopsys, Inc., Mountain View, 2019
143. Lorenz, J.K., Asenov, A., Baer, E., Barraud, S., Kluepfel, F., Millar, C., Nedjalkov, M.: Process variability for devices at and beyond the 7 nm node. ECS J. Solid State Sci. Technol. **7**, 595–601 (2018)
144. Wang, X., Reid, D., Wang, L., Burenkov, A., Millar, C., Lorenz, J., Asenov, A.: Hierarchical variability-aware compact models of 20 nm bulk CMOS. In: Goldsman, N., Stettler, M. (eds.) Proceedings SISPAD 2015, pp. 325–328. IEEE, Piscataway (2015)
145. Lorenz, J., Bär, E., Barraud, S., Brown, A., Evanschitzky, P., Klüpfel, F., Wang, L.: Process variability – technological challenge and design issue for nanoscale devices. Micromachines. **10**(1), 6 (2019). https://doi.org/10.3390/mi10010006

Jürgen Lorenz (Diplomas in Physics and Mathematics from TU Munich, Dr.-Ing. from University of Erlangen-Nuremberg, Germany) joined Fraunhofer in 1983. He has been in charge of the simulation activities at the now-called Fraunhofer IISB in Erlangen since 1985, where he is currently head its department “Modeling and Artificial Intelligence.” His main interests are physical models and software for process simulation.

Peter Pichler received his Dr.techn. degree from the TU Wien in Austria in 1985 and his Venia Legendi from the Friedrich-Alexander University Erlangen-Nürnberg in Germany in 2004. Since 1986 he has been Group Manager for Doping and Device Simulation at the Fraunhofer IISB in Erlangen, focused on dopant and point defect diffusion phenomena in semiconductors.