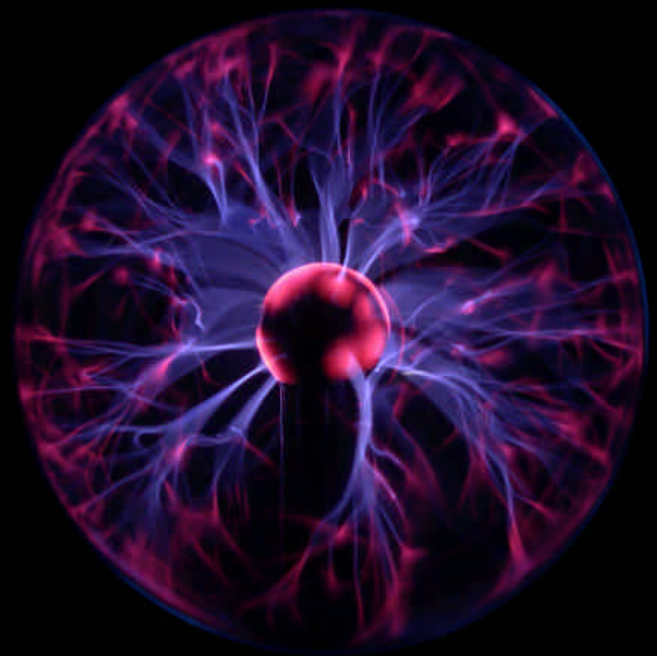
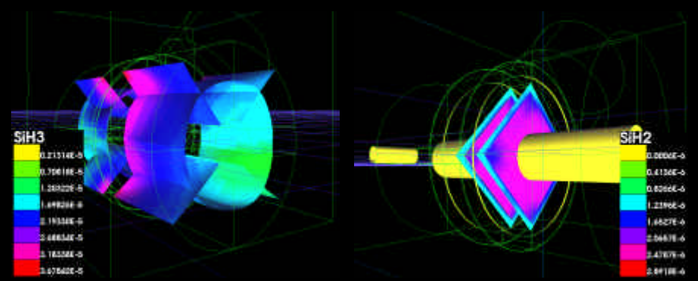


A Virtual Reactor for Simulation of Plasma Enhanced Chemical Vapor Deposition

Valeria Krzhizhanovskaya

A Virtual Reactor for Simulation of Plasma Enhanced Chemical Vapor Deposition



Valeria Krzhizhanovskaya

Invitation

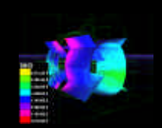
to attend the public
defense of my Ph.D. thesis

A Virtual Reactor for Simulation of Plasma Enhanced Chemical Vapor Deposition

Friday, 20 June 2008
12:00

Agnietenkapel

Oudezijds Voorburgwal 231
Amsterdam



There will be a reception
after the ceremony

Valeria Krzhizhanovskaya
valeria@science.uva.nl

A Virtual Reactor for Simulation of Plasma Enhanced Chemical Vapor Deposition

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Universiteit van Amsterdam
op gezag van de Rector Magnificus
prof. dr. D.C. van den Boom
ten overstaan van een door het college voor promoties ingestelde
commissie, in het openbaar te verdedigen in de Agnietenkapel
op Vrijdag 20 Juni 2008, te 12:00 uur

door

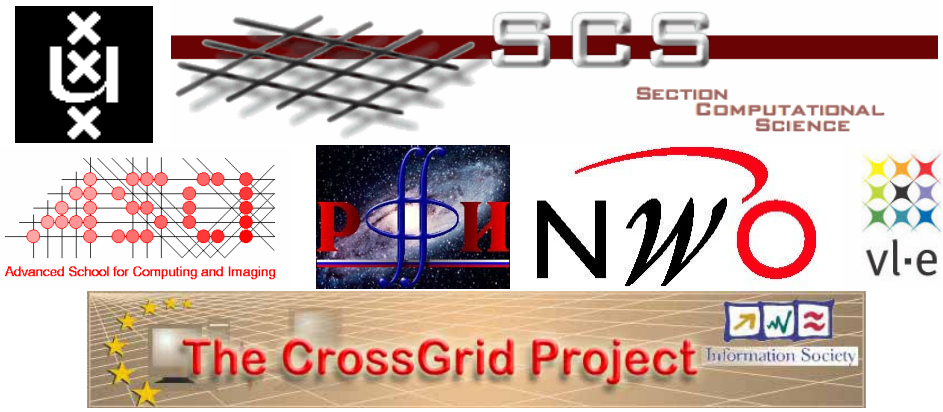
Valeria Vladimirovna Krzhizhanovskaya

geboren te St. Petersburg, Rusland

Promotiecommissie:

Promotor:	**Prof. Dr. P.M.A. Sloot**
Co-promotor:	**Prof. Dr. Yu.E. Gorbachev**
Overige leden:	Prof. Dr. S.E. Alexandrov
	Prof. Dr. M.T. Bubak
	Prof. Dr. W.J. Goedheer
	Dr. A.G. Hoekstra
	Dr. J.-P. Taran
	Prof. Dr. R.J. Meijer
Faculteit:	Faculteit der Natuurwetenschappen, Wiskunde en Informatica

The work described in this thesis has been carried out in the Section Computational Science of the University of Amsterdam and St. Petersburg State Polytechnic University, with the financial support of EU IST CrossGrid (IST200132243), Dutch Bsik Virtual Laboratory for eScience (VL-e), SP2.1, Dutch-Russian NWO-RFBR # 047.016.007, 047.016.018 and DAS Distributed ASCI Super Computer.



Author contact: valeria@science.uva.nl, lera@csa.ru

Cover design by Svetlana Krzhizhanovskaya
Back cover: plasma lamp by Luc Viatour from www.lucnix.be

ISBN: 978-90-9023166-2
Printed by PrintPartners Ipskamp, Enschede

To my family…

Contents

Chapter 1. Introduction

1.1. Motivation and scientific challenges

Plasma-Enhanced Chemical Vapor Deposition (PECVD) is nowadays a key sector of the industrial production of silicon-based films, an important material in modern microelectronics, notably transistors, computer chips and solar cells. This industrial sector stimulates research and development of the deposition processes and its physical chemistry. R&D efforts bear on many disciplines: fluid mechanics, chemistry gas-surface interactions to mention but a few. In this context, it must be noted that the engineering and optimization of reactors is much too costly and time-consuming to be left to the sole approach of trial and error, whereby various geometries are tested and then flow rates, temperatures and discharge currents adjusted while growth rates and surface quality are being monitored. Today, advances in numerical simulation offer great perspectives for aiding in the design of new reactors, optimizing the PECVD process and assistance in data reduction and interpretation, etc. The task is not easy, because so many processes need to be accurately taken into account and because expert usage of the most advanced numerical methods is necessary.

Modeling and simulation of the PECVD process is essentially a multidisciplinary endeavor and the state-of-the-art is far from satisfactory. Industry is using codes it has either purchased or developed on its own. Naturally, these codes are proprietary, and they are rarely available to the researchers in universities and research establishments. There also exist commercially-available codes; the latter offer limited flexibility, are not open and are often inefficient. Yet academic research laboratories would greatly benefit from these tools. Indeed, it is precisely in these places that some of the key expertise is found. They have the competence needed to study the complex chemical reactions and to recommend reduced schemes and critical paths, they understand gas-surface mechanisms and can propose and test new precursors. The applied mathematics laboratories are now in a position to cooperate with them fruitfully and to allow them to accomplish great progress, by offering advanced open tools of simulation and parallelization techniques. The work presented in this thesis was launched precisely along this line.

The goal we set out to reach was to develop models and codes for simulation of PECVD technology, to validate them, and to investigate the processes underlying this technology. The models have been validated using data from the literature and from experimental labs partners to the project, most notably Utrecht University, FOM Institute for Plasma Physics and Ioffe Institute of Technical Physics. The codes developed were in turn used to assist the latter in optimizing chamber design and process parameters.

In designing the codes, we followed the modern trend in modeling and simulation technology, departing from rigid simulation codes treating a fixed aspect, towards sophisticated Problem-Solving Environments (PSEs) [1]. These environments are widely

considered to be an essential emerging technology with high impact across all fields of science and engineering. PSEs are especially important to successfully address all kinds of multidisciplinary problems, such as multiphysics multiscale systems, fluid-structure interaction and rapidly developing fields like bio-informatics and other applications where vast amounts of data need to be managed and processed to discover patterns and knowledge contained within them. The research into understanding and improving the complex PECVD processes clearly benefits from such an environment.

The challenge addressed in this thesis is to develop a computationally efficient problem solving environment for PECVD: 'The Virtual Reactor', modeling all the essential multi-physics and multi-chemistry processes.

This work is one of the first attempts to treat 3D flow and 2/3D plasma discharge, thanks to the efficient parallel algorithms and the computational environment taking advantage of Grid computing technologies. This environment has allowed us to solve generic problems posed by modeling complexity and to build a flexible scalable high-performance *Virtual Reactor* for PECVD simulation.

1.2. Background

This subchapter provides a short overview of the PECVD processes, the related modeling and simulation, as well as the simulation environments. More details on the state-of-the-art in these areas can be found in the corresponding chapters.

1.2.1. Modeling and Simulation of PECVD processes

A typical PECVD reactor is a low-pressure chamber of a few liters volume, through which silane (SiH_4), often diluted with molecular hydrogen or argon, is pumped at a low flow rate. A gas mixture enters the reactor at room temperature. In the middle part of the chamber, two parallel electrodes are located (see Fig. 1.1). A substrate for film growth is located at the grounded electrode. To the other electrode, a radio-frequency (RF) oscillating voltage is applied.

A glowing capacitively coupled discharge between the electrodes initiates the ionization and dissociation processes. This leads to the formation of primary silane radicals, which in turn react with each other and with molecules to produce larger radicals. The radicals having free bonds attach to the surface of the substrate, thus composing the actual thin film. To enhance the chemical processes leading to the radicals formation, the substrate is heated to a temperature of around 250 ºC. The products of the chemical reactions that do not deposit are pumped away from the chamber. We should note that together with these products, an expensive silane source gas is also pumped out. One of the goals of the PECVD optimization is increasing the silane consumption ratio.

The quality of the growing films depends, among others, on the percentage of different species contributing to the growth and on the spatial homogeneity of deposition along the substrate. A deeper insight into these aspects of the deposition processes is another goal of the simulation.

A number of intertwined processes occur in a PECVD reactor: the plasma discharge processes (ionization, dissociation, excitation, recombination, attachment, ion

bombardment etc.), convective and diffusive transport, homogeneous (in the bulk) and heterogeneous (on the surfaces) chemical reactions and deposition, heat transfer, etc.

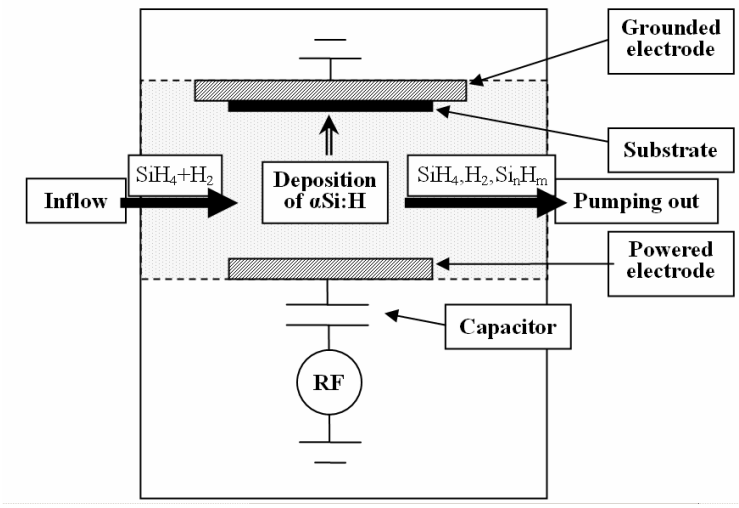


Fig. 1.1. Scheme of a PECVD reactor illustrating the basic principles. The shaded area indicates the discharge zone. In real reactors the geometry is much more complicated. Gases are allowed to flow through the total volume of the chamber

The modeling of these complicated processes can be simplified by noting that we can distinguish the discharge volume, where the plasma processes should be properly simulated, from the whole reactor volume, where only chemical reactions and transport processes should be computed (see Fig. 1.1). The rationale for this simplification stems from the two facts: (1) the energy of the electrons decreases rapidly outside the discharge, thus the high-energy electrons causing dissociation are available only in the discharge volume, and (2) the charged particles are neutralized via ion recombination on a very short distance away from the plasma discharge zone, and therefore can be neglected outside of the discharge volume. Based on these observations, we propose a hierarchical modeling approach with two distinct computational domains where different models and chemical processes shall be considered. Additional details on how this approach works are given in Chapter 2.

To simulate the multidimensional problems, several models have been developed: for 1D, 2D and 3D reactive flow dynamics, and 1D and axisymmetric 2D plasma discharge. Navier-Stokes equations were used to describe the 2D and 3D reactive flow; for the 1D case an original method was developed [2]; for the plasma we used Boltzmann equation together with Poisson equation and particle, momentum, and energy balance equations. For solving the resulting sets of stiff equations, different numerical approaches and algorithms were used.

1.2.2. Computational environment

Traditionally, research codes have been just pieces of computer programs without any user interface, only accessible to their developers, and not to the external user. The only "human" interface was the input and output data files, difficult to compose, understand and interpret. Commercial simulation packages, on the other hand, provide convenient user-friendly interfaces, but do not give enough flexibility in model formulation, choice of equations and numerical methods. One of the steps forward we accomplished was the move from rudimentary programs to easy-to-use research environments.

The importance of fully integrated PECVD simulators (Virtual Reactors) is recognized by various research groups and scientific software companies [3]. The complexity of the simulation necessitates an advanced user-friendly environment to explore the underlying physics and chemistry and associated design characteristics. In order to facilitate efficient data transport, analysis and archiving, as well as computationally efficient parameter-sweeps, we used a Grid-based PSE technology to provide an integrated simulation environment.

Connectivity between distant locations and interoperability between different kinds of systems and resources are some of the most promising characteristics of the Grid. Besides, the computational resources of the Grid provide the required computing power for large-scale simulations and complex visualization and collaborative environments, which are expected to become of major importance to many areas of computational physics.

Within the Virtual Reactor we merged the newly developed distributed interactive simulation and optimization environment [4] with results from the community for scientific computing [5,6] to create an architecture for Grid-enabled PSEs. The PSE considered here is generic in that it can handle High Performance Computing applications (Capability Computing: one parameter setting calculated as fast as possible) as well as High Throughput Computing (Capacity Computing: sweeping various parameter settings) applications.

The middleware we used to build the Virtual Reactor is based on the CrossGrid software system [7] and previous work on integrated environments for computational physics [8]. The CrossGrid pan-European distributed testbed shared resources across sixteen European sites, ranged from relatively small computing facilities in universities, to large computing centers, offering an ideal mixture to test the possibilities of an experimental Grid framework. National research networks and the multi-gigabit pan-European network Geant [9], assured interconnectivity between all sites. Experiments on optimization of load balancing for heterogeneous systems have been conducted on a Russian-Dutch RidGrid testbed.

1.3. Overview of the thesis

The thesis is organized as follows:

Chapter 2 describes the 1D reacting flow and 1D plasma models, discusses the discretization and implementation aspects, and presents the validation of our computational models against experimental data. It also demonstrates simulation results exploring the

influence of PECVD operating conditions on the deposition process and a comparison with analytical results.

Chapter 3 covers the 2D and 3D flow models and implementation, as well as simulation results with the 1D plasma model.

Chapter 4 describes the 2D plasma model and simulation results with the 3D flow.

Chapter 5 introduces the parallel algorithms and load balancing technique for heterogeneous Grid resources.

Chapter 6 details the Grid-based Virtual Reactor problem solving environment.

Chapter 7 brings a summary and conclusions; Chapter 8 gives a summary in Dutch; Chapter 9 offers acknowledgements and Chapter 10 lists the publications by the author.

1.4. References

1. http://www.npk.gov.pl/n/p/2/pse.html
2. Gorbachev, Yu.E., Zatevakhin, M.A., Krzhizhanovskaya, V.V., Schweigert, V.A., 2000, "Special Features of the Growth of Hydrogenated Amorphous Silicon in PECVD reactors," Tech. Phys., 45, N 8, pp.1032-1041.
3. http://www.cfdrc.com/bizareas/emergtech/plasma, http://www.fluent.com/about/news/pr/pr10.htm , http://www.semitech.us/consulting/pvt_growth/models, http://www.softimpact.ru/vr.php
4. Belleman, R.G. and Sloot, P.M.A., 2000, "The Design of Dynamic Exploration Environments for Computational Steering Simulations," Proc. of the SGI Users' Conference, pp. 57-74.
5. Marinescu, D.C. and Bölöni, L., 2000, "A component-based architecture for problem-solving environments," Math. Comp. Sim. 54, pp. 279-293.
6. Zhao, Z., Belleman, R.G., Van Albada G.D., Sloot, P.M.A., 2002, "AG-IVE: an Agent based solution to constructing Interactive Simulation Systems," LNCS, 2329, pp. 693-703.
7. CrossGrid EU Science project: http://www.eu-CrossGrid.org/
8. Iskra K.A., Belleman R.G., Van Albada G.D., Santoso J., Sloot P.M.A., Bal H.E., Spoelder H.J.W. and Bubak M., 2002, "The Polder Computing Environment, a system for interactive distributed simulation," Concurrency and Computation: Practice and Experience (Special Issue on Grid Computing Environments), 14, pp. 1313-1335.
9. Geant: http://www.dante.net/server.php?show=nav.007

Chapter 2. Modeling and Simulation Part I: 1D Flow and 1D Plasma Discharge*

2.1. Introduction

In this chapter, our goal is to develop a simplified model allowing one to describe the processes in a PECVD reactor and to predict some of the parameters measured in experiments. We are aware that a 1D model can give only limited agreement and trends. More accurate simulations are to come from 2D and 3D expanded models which are introduced in the subsequent chapters.

One of the most widespread methods of growing hydrogenated amorphous silicon films is the technology based on deposition from the gas phase of silicon-containing radicals produced by the decomposition of silane by RF discharge plasma in PECVD reactors. At present, various types of such systems are used. In one of these, the diode system, the substrate serves as one of the electrodes (see Fig. 1.1). In another system, belonging to a widely used class of systems with remote plasma and usually called a triode system, the substrate is placed outside the plasma discharge zone (see Fig. 2.1 and Section 2.7.7). This triode system offers the advantage of sheltering the substrate from ion bombardment, which can impair the surface quality. It was proposed and thoroughly investigated in [1]. By increasing the growth rate and film quality through optimizing the parameters of PECVD reactors, significant technological advantages can be attained. Therefore, various models are being proposed, which, with differing degrees of detail, describe the processes in the growth chamber and allow improvements to be made in the reactor design.

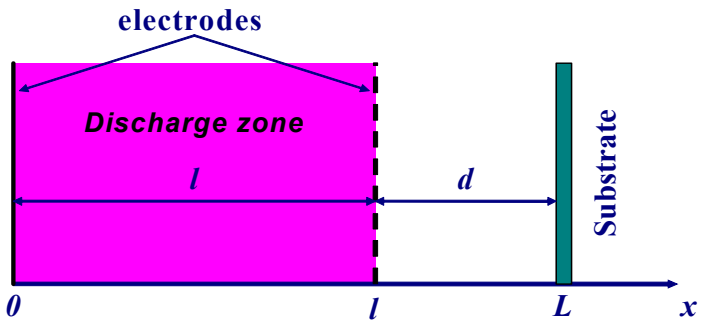


Fig. 2.1. Triode configuration and notations.

* Parts of this chapter were published in [12-15]

The first comprehensive model of the growth process of hydrogenated amorphous silicon film was proposed in [2].

A rather simple and effective model developed later [3] is suitable for making prompt estimates of all basic process parameters with sufficient accuracy. Apart from silane (SiH_4), 12 major chemical components of the mixture are included in this model: SiH_n , n = 1–3; H; H_2 ; Si_2H_n , n = 3–6; Si_3H_8 ; Si_2H_6* ; and Si_2H_6** (asterisks denote the excited electronic state). Recently, a similar model but with a different set of chemical components and a much more elaborate presentation of the electronic subsystem was proposed [4]. The authors carried out quite detailed measurements of the system parameters and provided the experimental data necessary for verifying the models.

In the framework of the model proposed in [3], the authors carried out a computational study of the physical and chemical processes occurring in the pure silane plasma in a diode reactor and performed a detailed analysis of the role of various components and the effect of the system parameters on film growth for major operating regimes. It was found that at low pressures molecular hydrogen accumulates in the chamber, so that silane ceases to be the only carrier component. In addition, a technology based on diluting silane with hydrogen is often used in practical applications, which saves silane and makes the production of silicon films ecologically cleaner. Under these conditions, the model of [3] is inadequate.

In this study, a model that takes into account the changes in silane concentration during film growth is presented. Transport due to diffusion was simulated using, instead of fixed coefficients, those calculated by Wilke's formula representing a first approximation of diffusion in a multicomponent medium. It has been shown that further elaboration of the description of diffusion processes does not improve the accuracy of the calculations of the component fluxes onto the film surface. Simuations of the plasma discharge were carried out using a fluid model [5] providing a means to calculate the electron densities.

Throughout this thesis, a modeling approach is used where the complete simulation domain is subdivided into two parts: the discharge zone, where the plasma processes should be properly simulated, and the rest of the reactor volume, where only chemical reactions and transport processes should be computed (see also Section 1.2.1 and Fig. 1.1). This subdivision is justified by the fact that the electron energy experiences a very rapid exponential decrease outside the discharge, therefore electrons do not cause dissociation there. In addition, electron-ion recombination eliminates the charged particles already at a very short distance from the plasma discharge zone, therefore only the neutrals shall be considered there [12,16,17]. This effect will be illustrated in Section 4.3 where the results of 2D plasma simulation are presented. In the discharge volume, a combined kinetic-fluid plasma model is used (described in the next Section), and in the whole reactor volume a gas dynamic approach is used for the reactive flow simulation (described in Section 2.3). The plasma simulation provides the reactive flow code with the electron density distribution averaged over one RF cycle, while the plasma model uses the background gases concentrations provided by the flow code. The process of data exchange is organized iteratively, and proceeds until the background gas concentration does not change any longer.

This approach is not entirely self-consistent. Ideally, one should simulate plasma and flow with full coupling, solving the whole set of plasma and flow equations together. However, that would be computationally prohibitive because of the large gap in scales (several orders of magnitude) between the required time steps for plasma and flow. Other limitations of this approach are the following: it is assumed that ions do not contribute to the film growth (which can add about 10%), and all the electrons inside the discharge zone are assumed to have sufficient energy to cause dissociation.

In this Chapter, expressions for the rate constants of reactions initiated by electron impact in silane-hydrogen mixtures have been derived. Hence, the processes in reactors could be studied for a wider range of parameters, and the processes taking place in silane diluted with molecular hydrogen could be simulated. The proposed model describes the processes in triode-type reactors with the discharge zone moved away from the substrate. Finally, in this work, the effect of the reactor volume, which in almost every real installation does not coincide with the discharge zone, has been accounted for in a correct manner.

2.2. 1D plasma model

Direct Monte Carlo simulation of electron kinetics showed that at pressures higher than 0.1 Torr we can use a gas-dynamic approach for electron and ion description. In the present work we used a 1D fluid model [5] for simulation of the RF discharge processes. The continuity equations for electrons, ions and neutrals are solved consistently with the balance equation for the average electron energy, and with the Poisson equation for the electric field distribution. The model accounts for the emission and action of secondary electrons produced by the primary electron impact on the surface of the electrode.

The rates of inelastic electron collisions and the electron transport coefficients are calculated by solving the Boltzmann equation for the electron energy distribution function (EEDF) f_0 in the two-term approximation: [5]

$$-\frac{\partial}{\partial\varepsilon}\left(\frac{2e^2E^2\varepsilon}{3mN\sigma_t}\frac{\partial f_0}{\partial\varepsilon}\right)=St\{f_0\} \tag{2.1}$$

Here, ε is electron energy, N is neutral gas number density, E is electric field, σ_t is scattering cross-section, m is electron mass, and collision term $St\{f_0\}$ includes only inelastic electron-ion collisions.

The EEDF is calculated as a function of the electric field for a given composition of the neutral background density. A look-up table is then constructed to obtain the collision rates and electron transport coefficients used in the fluid model as functions of the average electron energy.

While calculating the EEDF, the electric field was assumed to be spatially uniform and constant in time. This approximation was compensated for in the fluid model by the spatial and temporal variation of the averaged electron energy.

Further details on this RF discharge model have been published in [5], including the list of all chemical reactions simulated in the plasma discharge volume (ionization, dissociation, excitation, recombination, attachment).

Within the fluid model, we solve the continuity and average electron energy equations for electrons (denoted with subscript *e*), and continuity and momentum equations for positive and negative ions (denoted with *p* and *n* subscripts):

$$\begin{cases} \dfrac{\partial N_e}{\partial t} - \dfrac{\partial\left(\mu_e E N_e + \partial\left(D_e N_e\right)/\partial x\right)}{\partial x} = \left(\nu_i - \nu_a\right) N_e + \nu_d N_n - \nu_b N_e N_p \\ \dfrac{\partial N_e U_e}{\partial t} - \dfrac{\partial\left(\mu_u E N_e U_e + \partial\left(D_u N_e U_e\right)/\partial x\right)}{\partial x} = eE\left(\mu_e E N_e + \dfrac{\partial D_e N_e}{\partial x}\right) - \nu_u N_e U_e \end{cases} \tag{2.2}$$

$$\begin{cases} \dfrac{\partial V_p}{\partial t} + \dfrac{\partial V_p^2/2}{\partial x} = \dfrac{eE}{M_p} - W_p N V_p \\ \dfrac{\partial V_n}{\partial t} + \dfrac{\partial V_n^2/2}{\partial x} = \dfrac{-eE}{M_n} - W_n N V_n \end{cases} \tag{2.3}$$

$$\begin{cases} \dfrac{\partial N_p}{\partial t} + \dfrac{\partial N_p V_p}{\partial x} = \nu_i N N_e - \nu_b N_e N_p - \beta N_p N_n \\ \dfrac{\partial N_n}{\partial t} + \dfrac{\partial N_n V_n}{\partial x} = \nu_a N N_e - \nu_d N_n - \beta N_p N_n \end{cases} \tag{2.4}$$

Here t is time, x is position along axis X perpendicular to the electrodes, V_n , V_p are velocities of negative and positive ions; M_n , M_p, W_n , W_p are masses and frequencies of ions scattering; β is ion-ion recombination rate; e is elementary charge; E is electric field; N, N_e , N_n , N_p are number densities of neutrals, electrons, negative and positive ions; ν_i , ν_a , ν_d , ν_u are frequencies of ionization, attachment, detachment and inelastic collisions of electrons; μ_e , D_e are electron mobility and diffusion transport coefficients; μ_u , D_u are mobility and diffusion coefficients for the average electron energy U_e; and ν_b is electron-ion recombination rate.

The set of equations (2.2-2.4) are solved consistently with the Poisson equation for the potential Φ:

$$\frac{d^2\Phi}{dx^2} = 4\pi e\left(N_e + N_n - N_p\right), \quad E = -\frac{d\Phi}{dx}\,. \tag{2.5}$$

The secondary electrons, also called γ-electrons, emitted from the electrode surface can gain very high energy (of order of plasma potential, 10-100 eV), and thus can initiate a number of ionization actions. The usual gas-dynamic approach does not take that into account, but averages the electron energy over all electrons (most of them have an energy of order of electron temperature, about 1-3 eV). Therefore the second type of electrons needs to be introduced, for which we solve equations (2.2) separately.

Boundary conditions take into account ion-electron emission:

$$\begin{cases} \dfrac{N_e V_t}{4} - \mu_e E N_e - \dfrac{\partial D_e N_e}{\partial x} = k_{em} N_p V_p \\ \dfrac{N_e V_t U_e}{4} - \mu_u E N_e U_e - \dfrac{\partial D_u N_e U_e}{\partial x} = k_{em} N_p V_p U_e \end{cases} \tag{2.6}$$

Here V_t denotes thermal velocity of electrons and k_{em} ion-electron emission coefficient.

To account for gas heating, a heat transfer equation was solved:

$$\frac{d\lambda(dT/dx)}{dx} = Q \tag{2.7}$$

with λ, thermal conductivity coefficient; Q, the Joule heating due to ion and electron energy loss and T, gas temperature.

The pressure $p = NkT$ is set constant.

The data on chemistry, kinetic coefficients, cross-sections and other model details can be found in [5].

Special attention was paid to resolving the sheaths, an important zone within the discharge volume where the electric field and other plasma parameters, like electron and ion concentration and energies, are different from those in the bulk. To resolve the sheath, the computational mesh was manually refined near the electrode walls.

Validation and extensive testing of the model described above has been performed earlier [5]. The main plasma parameters predicted by simulations were found to be in good agreement with experimental data. We additionally validated the model in the parameter range used in our numerical experiments, and later compared the results of our simulations with those calculated by another 1D model developed in the group of Prof. Goedheer [4]. The electron concentration profiles (spatial distribution between the electrodes) produced by the two models slightly differed in the sheath width (see Fig. 2.2), however the net effects (the mean electron concentration and the film growth parameters) simulated with two models were reasonably close to each other.

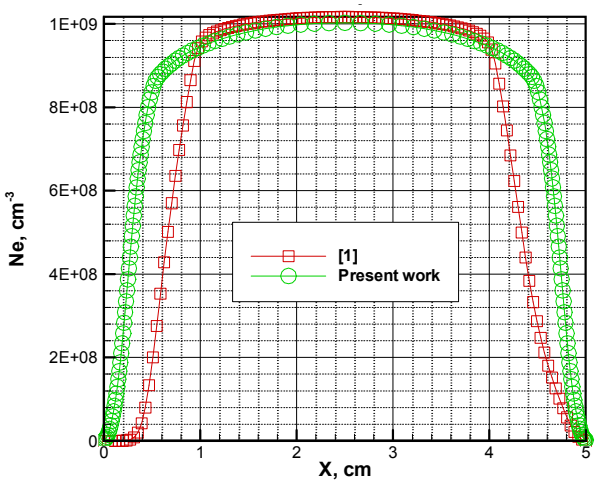


Fig. 2.2. Comparison of electron distribution predicted by the two models. p=0.15Torr, W=0.0366 W/cm^2. In the figure, [1] indicates the result predicted by the 1D model developed in the group of Prof. Goedheer [4].

2.3. 1D reactive flow dynamics model

In the proposed model of a PECVD reactor, the gas-phase processes in the chamber are described by one-dimensional equations of chemical kinetics, in which, as estimated in [3], only the diffusion transfer processes are taken into account:

$$\frac{\partial n_i}{\partial t} = \frac{\partial}{\partial x}\left(D_i \frac{\partial n_i}{\partial x} \right) + F_i \,. \tag{2.8}$$

Here, n_i is the number density of the i-th component and F_i is the source term due to chemical reactions (more details can be found in [3]). The effective diffusion coefficient of the i-th component D_i is calculated by Wilke's formula [6], commonly used in practical calculations of multicomponent gas mixtures:

$$D_i = (n - n_i) \Big/ \sum_{k=1,k\neq i}^{Nk} \frac{n_k}{D_{ki}} \text{, cm}^2\text{/s} \tag{2.9}$$

where n is the total number density of the mixture.

The binary diffusion coefficient D_{ki} is calculated by the formula of the molecular kinetic theory of gases [7] with use of the Lennard-Jones 6-12 potential:

$$D_{ki} = 0.002628 \cdot \frac{\sqrt{T^3/2\mu_{ki}}}{p\sigma_{ki}^2 \Omega_{ki}^{(1.1)*}\left(T_{ki}^*\right)} \text{, cm}^2\text{/s} \tag{2.10}$$

Here, $\mu_{ki} = m_k m_i/(m_k + m_i)$ is the effective mass of the kth and ith species, m_i is the molecular weight of the ith species, $\sigma_{ki}=(\sigma_k+\sigma_i)/2$ is the effective collision diameter, $T_{ki}^*= kT/\varepsilon_{ki}$ is the characteristic temperature, $\varepsilon_{ki} =(\varepsilon_k\varepsilon_i)^{1/2}$ is a parameter of the potential for intermolecular interaction (potential well depth), k is the Boltzmann constant, and $\Omega_{ki}^{(1.1)*}(T_{ki}^*)$ is a reduced Ω-integral of collisions for mass transfer normalized by the Ω-integral of the model of hard spheres. Values of the T_{ki}^* function for a wide range of characteristic temperatures T_{ki}^* can be found in monograph [7]; in the calculations, approximate formulas given in [8] were used. Values of the Lennard–Jones potential parameters σ_i and ε_i borrowed from [9] are given in Table 2-1.

Table 2-1. Parameters of the Lennard–Jones potential

Component	σ_i, Å	ε_i/k, K
SiH_4	4.084	207.6
SiH_3	3.943	170.3
SiH_2	3.803	133.1
SiH	3.662	95.8
H	2.5	30.0
H_2	2.915	59.7
Si_2H_6	4.828	301.3

Component	σ_i, Å	ε_i/k, K
$Si_2H_6^*$	4.828	301.3
$Si_2H_6^{**}$	4.828	301.3
Si_2H_5	4.717	306.9
Si_2H_4	4.601	312.6
Si_2H_3	4.494	318.2
Si_3H_8	5.562	331.2

In [10], experimental determinations of the ambipolar diffusion coefficients of silyl in silane and molecular hydrogen were carried out and found to be equal to 140±30 and 580±140 cm^2/s, respectively, at the temperature 320 K and pressure 1 Torr. Under these

conditions, formula (2.10) gives the values of 111 and 536 cm^2/s, respectively, which is in good agreement with the measured data.

In the next section we describe the boundary conditions that take into account the deposition process.

2.4. Modeling chemical processes and deposition

2.4.1. Model of the deposition process

To model the deposition process, a new model of sticking probabilities was developed [12]. In this model only neutral radicals having free bonds contribute to the film growth (Si_nH_m for *m=2n-1, m=2n, m=2n+1, n≥1*). Neither detachment nor ion bombardment is considered, which potentially can influence the deposition process and film properties and should be included in the model in the future, as was done in [18]. There is a possibility to further improve the model by refining the experimental data that allowed estimating the sticking probabilities; yet a better way would be to develop an atomic-level model for simulation of heterogeneous reactions on the surface. The latter approach is something for the future, since it will definitely require extensive experimental and theoretical developments.

To implement the deposition model, a boundary condition with a concentration jump near the surface was introduced. The complete model of surface processes is based on a simplified kinetic theory which allowed us to derive the boundary conditions for the species concentration jump [3]:

$$-D_i \frac{\partial n_i}{\partial x} = \frac{s_i c_i}{2(2-s_i)} n_i \tag{2.11}$$

Here, $\partial/\partial x$ is a normal derivative at the outside of the surface, $c_i = 2\sqrt{2kT/\pi m_i}$ is the thermal velocity, and s_i is the sticking probability of the *i*-th species. Sticking probabilities are chosen as follows [2,3]: $s_i = 0$ for Si_nH_{2n+2} (no free bonds), $s_i = 0.15$ for Si_nH_{2n+1} (1 free bond), $s_i = 1$ for Si_nH_{2n}, Si_nH_{2n-1} (2 free bonds). In [3] it was shown that the actual values of the sticking coefficients s_i do not significantly influence the deposition rate.

The deposition rate is calculated as follows:

$$R = -v_{Si} \sum_{i=1}^{N} q_i D_i \frac{\partial n_i}{\partial x}, \tag{2.12}$$

where $v_{Si} = 2.195 \cdot 10^{-29}\ m^3$ is the silicon atom volume, and q_i is the number of silicon atoms in a depositing radical.

In this simplified model, we did not consider surface reactions leading to hydrogen release from the surface, which should be taken into account for accurate predictions of the film composition. We shall note however that this omission will not affect the net growth rate (calculated from the silicon atom contents), thus the model is well suitable for the film deposition rate estimations. Hydrogen concentration in the film can be corrected within our approach by comparing the simulation results with experimental data and finding reaction rate constants of the missing surface reactions. This procedure will require a large collection of experimental data in a complete range of process conditions, and a separate

research that will allow identifying the most essential surface process and their rate constants. This will be one of the directions of future work.

2.4.2. Chemical components and reactions

Apart from silane (SiH_4) and diluent hydrogen (H_2), 12 contributing radicals are included in this model: H, SiH_n , n = 1–3; Si_2H_n , n = 3–6; Si_3H_8 ; $Si_2H_6^*$; and $Si_2H_6^{**}$ (asterisks denote the excited electronic state).

Chemical reactions considered essential for adequate modeling are listed in **Table 2-2**. In addition, silane and disilane pyrolysis reactions are taken into consideration, which, as is indicated by the simulations, give an insignificant contribution to the total chemical balance.

Table 2-2. List of chemical reactions and reaction constants used in this Chapter.

N^o	Reactions	Reference rate constants, cm^3/s
R1	$SiH_4 + e \rightarrow SiH_3 + H + e$	$3.000 \cdot 10^{-11}$
R2	$SiH_4 + e \rightarrow SiH_2 + 2H + e$	$1.500 \cdot 10^{-10}$
R3	$SiH_4 + e \rightarrow SiH + H + H_2 + e$	$9.340 \cdot 10^{-12}$
R4	$SiH_4 + e \rightarrow SiH_2 + H_2 + e$	$7.190 \cdot 10^{-12}$
R5	$H_2 + e \rightarrow 2H + e$	$4.490 \cdot 10^{-12}$
R6	$Si_2H_6 + e \rightarrow SiH_4 + SiH_2 + e$	$3.720 \cdot 10^{-10}$
R7	$Si_2H_6 + e \rightarrow Si_2H_4 + H_2 + e$	$3.700 \cdot 10^{-11}$
R8	$SiH_4 + H \rightarrow SiH_3 + H_2$	$2.530 \cdot 10^{-12}$
R9	$SiH_4 + SiH_2 \rightarrow Si_2H_6^*$	$1.000 \cdot 10^{-11}$
R10	$SiH_4 + SiH \rightarrow Si_2H_3 + H_2$	$1.700 \cdot 10^{-12}$
R11	$SiH_4 + SiH \rightarrow Si_2H_5$	$2.500 \cdot 10^{-12}$
R12	$SiH_4 + Si_2H_5 \rightarrow SiH_3 + Si_2H_6$	$5.000 \cdot 10^{-13}$
R13	$SiH_4 + Si_2H_4 \rightarrow Si_3H_8$	$1.000 \cdot 10^{-10}$
R14	$SiH_3 + H \rightarrow SiH_2 + H_2$	$1.000 \cdot 10^{-10}$
R15	$SiH_3 + SiH_3 \rightarrow SiH_4 + SiH_2$	$1.500 \cdot 10^{-10}$
R16	$SiH_3 + SiH_3 \rightarrow Si_2H_6^{**}$	$1.000 \cdot 10^{-11}$
R17	$SiH_3 + Si_2H_5 \rightarrow SiH_4 + Si_2H_4$	$1.000 \cdot 10^{-10}$
R18	$SiH_3 + Si_2H_5 \rightarrow Si_3H_8$	$1.000 \cdot 10^{-11}$
R19	$SiH_3 + Si_2H_6 \rightarrow SiH_4 + Si_2H_5$	$3.270 \cdot 10^{-12}$
R20	$SiH_2 + H \rightarrow SiH + H_2$	$7.960 \cdot 10^{-13}$
R21	$SiH_2 + Si_2H_6 \rightarrow Si_3H_8$	$1.200 \cdot 10^{-10}$
R22	$Si_2H_3 + H_2 \rightarrow Si_2H_5$	$1.700 \cdot 10^{-12}$
R23	$Si_2H_4 + H_2 \rightarrow SiH_4 + SiH_2$	$1.000 \cdot 10^{-10}$
R24	$Si_2H_5 + H \rightarrow Si_2H_4 + H_2$	$1.000 \cdot 10^{-10}$
R25	$Si_2H_6 + H \rightarrow SiH_4 + SiH_3$	$7.160 \cdot 10^{-12}$
R26	$Si_2H_6 + H \rightarrow Si_2H_5 + H_2$	$1.430 \cdot 10^{-11}$
R27	$Si_2H_6^* \rightarrow Si_2H_4 + H_2$	$5.000 \cdot 10^{6}$ s^{-1}

R28	$Si_2H_6^* + M \rightarrow Si_2H_6 + M$ (M – any particle colliding with $Si_2H_6^*$)	$1.000 \cdot 10^{-10}$
R29	$Si_2H_6^{**} \rightarrow SiH_4 + SiH_2$	$2.300 \cdot 10^{7}$ s^{-1}
R30	$Si_2H_6^{**} \rightarrow Si_2H_4 + H_2$	$2.300 \cdot 10^{7}$ s^{-1}
R31	$Si_3H_8 + H \rightarrow Si_2H_5 + SiH_4$	$2.170 \cdot 10^{-11}$
R32	$Si_3H_8 + SiH_3 \rightarrow Si_4H_9 + H_2$	$1.000 \cdot 10^{-11}$

Data are given for the inter-electrode separation l^0 = 2.5 cm, partial pressures of molecular hydrogen and silane p^0 = 0.125 Torr, temperature T^0 = 520 K, and discharge power W^0 = 0.25 mW/cm^3.

2.4.3. Electron impact reaction rate constants

The approach used for calculating the electron-induced reaction rates, avoiding calculation of the corresponding electron temperatures, was proposed in [3] for pure silane plasmas and extended for diluted silane mixtures in [12]. This approximate method for determining the dependencies of the constants of electron-impact processes on the discharge parameters was based on a number of assumptions: the concentration of the background gasses (SiH_4 and H_2) was assumed to be much higher than the concentration of other species; the electron temperature was assumed to be sufficiently small compared to the ionization threshold; the plasma was considered quasi-neutral and the recombination losses small compared to the diffusion losses. The electron energy distribution function was modeled in the Boltzmann form, taking into account that the electron temperature corresponds to the high-energy "tail" of the electron distribution and depends on the pressure of the gas and the inter-electrode spacing. The parameterization formula obtained in [10] is applicable to the relatively low power discharges, where the contribution of the high-energy γ-electrons into ionization can be neglected. The applicability of this formula for high-power discharges remains to be validated against a comprehensive plasma simulation. The processes neglected while deriving this relationship (electron-ion recombination and γ-electrons emission) are properly taken into account in the plasma discharge model.

To define the dependence of the electron-induced reaction rate constants on partial pressures and the inter-electrode separation l, the following relationship between the reaction constants k_r and the ionization constants k_i was used [3]:

$$k_r / k_r^0 = \left(k_i / k_i^0\right)^{\varepsilon_r/\varepsilon_i} \tag{2.13}$$

where $\varepsilon_r, \varepsilon_i$ are the thresholds for the chemical reaction and ionization, respectively [11], and the superscript "0" refers to the reference conditions for which the values of the constants are known (see **Table** 2-2).

Then, assuming that the plasma as a whole is neutral, that the removal rate of electrons is equal to that of ions (mainly SiH_2^+ ions), and that the carrier mixture consists of silane and molecular hydrogen, the following expression can be obtained using formula (2.9):

$$\frac{k_i}{k_i^0} = \frac{n}{n^0}\frac{n^0_{SiH_4}}{n_{SiH_4}}\frac{\left(n^0_{SiH_4}/D^0_{SiH_4SiH_2} + n^0_{H_2}/D^0_{H_2SiH_2}\right)}{\left(n_{SiH_4}/D_{SiH_4SiH_2} + n_{H_2}/D_{H_2SiH_2}\right)}\left(\frac{l^0}{l}\right)^2 \tag{2.14}$$

Here, n and n^0 are the number densities of the species under present and reference conditions respectively, D_{ij} is the binary diffusion coefficient. For the reference conditions, a silane–hydrogen mixture is chosen with partial pressures of hydrogen and silane $p^0_{H_2} = p^0_{SiH_4} = 0.125$ Torr at the temperature $T^0 = 520$ K and the inter-electrode distance $l^0 = 2.5$ cm, because the constants for these conditions were determined in [2].

Comparison of the electron-induced reaction rate constants calculated with the expression (2.13)-(2.14) with those calculated from EEDF showed a good agreement in the parameter range used in our simulations.

2.5. Numerical methods, implementation and computational environment

2.5.1. Numerical methods and algorithms

The mathematical model describing the diffusion processes in a chemically reacting mixture is represented by a stiff system of nonlinear equations. For solving this system of equations (2.8), a numerical method that takes into account the specific features of these equations was developed [3]. The source terms F_i can be rewritten in the form of $F_i = f_i - A_i n_i$, with $f_i \geq 0$ being the chemical production rate of the ith species, $\partial f_i / \partial n_i = 0$, and $A_i \geq 0$ is the depletion rate. Then the following implicit numerical scheme can be used for each species:

$$\left(1 + \Delta t\, A_i^k\right) n_i^{k+1} - \Delta t\, D_i n_{i_{xx}}^{k+1} = n_i^k + \Delta t\, f_i^k , \tag{2.15}$$

where Δt is the time step value, superscript k is time step index, and $n_{i_{xx}}^{k+1}$ is the second spatial derivative of the i-th species concentration. Each time step, the system (2.15) is solved by the sweep-time algorithm. This approximation ensures a positive unconditional convergence of the algorithm for any value of the time step. An automatic time step selection procedure was used, based on controlling the maximal changes of relative concentrations per one time step. Extensive testing of the scheme showed that it allows to obtain stable convergence for $\Delta t >> \Delta t_{chem.}$, where Δt_{chem} is the characteristic time of the fastest chemical reaction. The simulation proceeds for the time τ during which the complete volume of the gas in the chamber will be fully replaced by the fresh gas supplied through the inlet: $\tau = M_0/Q$, where M_0 is the total mass of the gas in the chamber and Q is mass inflow rate. In real-life reactor chambers this characteristic "forced convection" time is much longer than the characteristic time of diffusion and chemical reactions, therefore within the time τ the system will arrive at a quasi steady state.

2.5.2. Implementation and computational environment components

The plasma and the 1D reactive flow models were implemented as FORTRAN computer codes. The coupling between the codes described in Section 2.1 consists in exchanging data files that contain the variable distributions. The codes are platform-independent and were compiled and used on Unix-like and Windows operating systems with the open-source GNU FORTRAN compilers. To facilitate the problem formulation and parameter setting procedure, as well as to organize parameter sweep simulations, a computational environment called REAF was developed, which served as a prototype of the Virtual Reactor environment developed and extended later [20].

2.6. Analytical results

With some assumptions it is possible to solve equations (2.8) analytically. That will allow fast estimation of tendencies and provide a fairly rigorous check of our numerical simulations for some test cases. Asymptotic analysis based on the difference in scale of various processes led to the analytical formulas explaining the influence of the chamber conditions on the deposition process. In [3], such relationships were derived for the diode system, and in this work, they were expanded for the triode configuration (see Fig. 2.1). The entire simulation domain of $0<x<L$ was divided into two subdomains: $0<x<l$, where the density n_e of electrons initiating the chemical reactions was taken equal to that in the inter-electrode gap, and $l<x<L$, where the electron energy is not sufficient to initiate chemical reactions, thus can be set to zero (see also explanations in section 1.2.1). It is assumed that in the latter zone of size $d=L–l$, only monomolecular reactions or reactions involving molecular collisions are taking place.

Estimates of basic processes similar to those in [3] produced the following simplified equation for the behavior of silyl, the main species contributing to the film growth:

$$D_{SiH_3}\, d^2 n_{SiH_3} \big/ dx^2 = 2k_8 n_H n_{SiH_4} - k_{32} n_{SiH_3} n_{Si_3H_8}, \tag{2.16}$$

where k_i is the constant of the i-th reaction (i is the reaction number in Table 2-2). The boundary condition (2.11) with a sticking coefficient $s_{SiH3}= 0.15$ is used.

A corresponding analysis for atomic hydrogen, which is the major initiator of silyl production, leads to the following balance equation in the discharge zone $0<x<l$:

$$-D_H\, d^2 n_H \big/ dx^2 = 2k_2 n_e n_{SiH_4} - k_8 n_{SiH_4} n_H\,. \tag{2.17}$$

Outside the discharge zone at $l<x<L$ the behavior of atomic hydrogen is described by the same equation, except that the first term on the right-hand side should be omitted ($n_e = 0$), which means that hydrogen escapes the discharge zone only through diffusion. Solving these equations with boundary conditions (2.11) in the absence of deposition ($s_H = 0$) and with the condition of continuity of solutions at the boundary between the zones (at $x=l$), we obtain an analytical expression for the atomic hydrogen concentration profile:

$$n_H = \frac{2k_2 n_e}{k_8}\left(1 - \frac{\sinh\left(d/L_H\right)}{\sinh\left(L/L_H\right)}\right)\cosh\frac{x}{L_H},\ 0 < x < 1\,, \tag{2.18}$$

$$n_H = \frac{2k_2 n_e}{k_8} \frac{\sinh\left(l/L_H\right)}{\sinh\left(L/L_H\right)} \cosh\frac{L-x}{L_H},\ l < x < L\ . \tag{2.19}$$

Substitution of expressions (2.18) and (2.19) into equation (2.16) yields an analytical expression for the flux of silyl towards the surface as a function of the system parameters:

$$\Gamma_{SiH_3} = \frac{2k_2 n_e n_{SiH_4} L_3}{\sinh\left(L/L_3\right)} \left[\begin{array}{l} 2\sinh^2\frac{l}{2L_3} + \frac{L_H^2}{L_3^2 - L_H^2} \frac{1}{\sinh\left(L/L_H\right)} \times \\ \left(\cosh\frac{l}{L_3}\sinh\frac{L}{L_H} - \cosh\frac{L}{L_3}\sinh\frac{l}{L_H} - \sinh\frac{d}{L_H} \right) \end{array} \right] \tag{2.20}$$

$$L_H = \sqrt{D_H/(k_8 n_{SiH_4})}\ ,\ L_3 = \sqrt{D_{SiH_3}/\left(k_{32} n_{Si_3H_8}\right)}\ . \tag{2.21}$$

Here, L_H and L_3 are reaction-limited diffusion lengths for atomic hydrogen and silyl, respectively, introduced in [3] and defined as the distance which a diffusing species will travel before undergoing transformation in chemical reaction. More details on diffusion-reaction lengths and their analysis for different species will be given in Section 2.7.8. For $l \to L$ ($d \to 0$), expressions (2.18)-(2.20) describe a diode system.

To conclude this section, we consider the behavior of silylyl. In the discharge zone, silylyl is produced mainly via the dissociation of silane by electron impact, reaction R2 in Table 2-2, and through recombination reaction R23, whereas, outside this zone, it is produced in recombination reaction R15 and, again, R23. Chemical removal of silylyl in both cases depends on its addition reaction with silane (R9). The Si_2H_4 radical is an intermediate product of the cycle of fast chemical reactions R9, R27, and R23; therefore its concentration is close to the equilibrium value $n^e_{Si_2H_4} = k_9 n_{SiH_2} n_{SiH_4} / (k_{23} n_{H_2} + k_{13} n_{SiH_4})$ [3]. If it were not for reaction R13 that diverts part of the silicon atoms to higher silanes, reactions R23 and R9 would compensate one another and the only mechanism for silylyl to escape from the discharge zone would be its diffusion to the surface for deposition and pumping out. Given the above considerations, we can derive the effective rate constant of chemical removal of silylyl from the discharge:

$$k_{eff} = k_9 k_{13} n_{SiH_4} / \left(k_{23} n_{H_2} + k_{13} n_{SiH_4}\right) \cong k_9 / \left(n_{H_2}/n_{SiH_4} + 1\right). \tag{2.22}$$

The last approximate equality is a consequence of $k_{13} \cong k_{23}$. Therefore, the behavior of silylyl can be described by the equations

$$-D_{SiH_2}\, d^2 n_{SiH_2} / dx^2 = 2k_2 n_e n_{SiH_4} - k_{ef} n_{SiH_2} n_{SiH_4},\ 0 < x < l \tag{2.23}$$

$$-D_{SiH_2}\, d^2 n_{SiH_2} / dx^2 = k_{15} n^2_{SiH_3} - k_{ef} n_{SiH_2} n_{SiH_4},\ l < x < L\ . \tag{2.24}$$

By solving these equations with corresponding boundary conditions, an analytical expression for the flux of silylyl (SiH_2) to the surface has been obtained, but, in view of its complexity, we do not present it here.

2.7. Simulation results

To validate the model we simulated the deposition process in a wide range of operating conditions and compared the simulation results with experimental data available.

Furthermore, we studied the influence of various controllable parameters on the processes occurring in a PECVD reactor. In our simulations, we paid special attention to parameters considered to be important for either adjusting or assessing the reactor performance, namely the influence of flow rate, reactor volume, silane dilution, substrate separation in remote plasma configuration, as well as the "dust" production.

2.7.1. Comparison with experiment

Comparison with experimental data in the literature is difficult, because the set of parameters describing the system available is far from complete. For this reason, we chose the results given in [4] for comparison. In this work, extensive information on the experimental conditions and results of numerical calculations are described.

The experimental installation was a chamber with a volume of 10 liter and contained two electrodes of radius 8 cm located 2.7 cm apart. For this configuration, the ratio of the reaction volume to the total volume of the chamber is $R_v = 0.054$ (for more details, see Section 2.7.5). The experiments were carried out with a mixture of silane and hydrogen in equal molar concentrations with a flow rate equal to 3.6 l/h under normal atmospheric conditions; the temperature in the reactor was 400 K. A program for computing parameters of the RF discharge between two plane-parallel electrodes was used to calculate the electron densities [5]. The main problem is to obtain a good estimate of the discharge power. In [4], from a comparison of the self-bias potential computed in simulations of argon plasma based on corresponding experiments, it was determined that the power directed to the discharge amounted to 50% of the total power. The same conclusion was made from a comparison of the calculated concentrations of H_2 and SiH_4 in the hydrogen–silane plasma with experimental data. Therefore, in calculations performed for the suggested model, we used a discharge power of 5 W, as has been proposed in [4].

Fig. 2.3 shows the dependence of the film growth rate on the discharge frequency at a constant pressure of 16 Pa in the reactor. As can be seen from this graph, the results are in good agreement with experimental data up to a frequency of ~ 30 MHz. At frequencies higher than ~ 30 MHz, the calculated curve tends to level off. The same tendency, although more gradual, was noted in the simulations reported in [4]. At the same time, in [4], the *experimental* dependence of the growth rate on the frequency was shown to be approximately linear. This discrepancy between experiment and theory seems to be persistent and is not well understood.

From the calculations of the partial pressures of molecular hydrogen and silane as functions of the discharge frequency given in Fig. 2.4, one can see that the data obtained using the approach developed in this study are in better agreement with experiment than the calculation results obtained in [4].

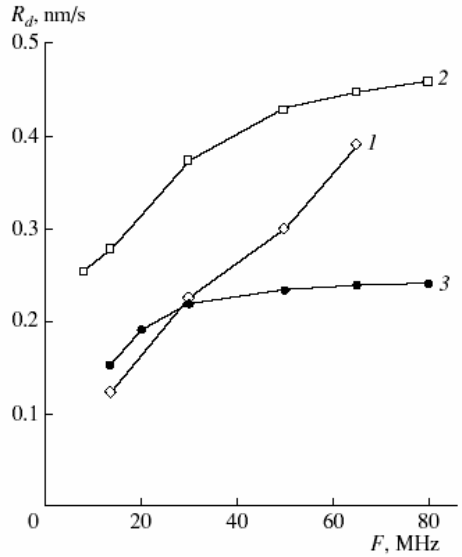


Fig. 2.3. Film growth rate as a function of discharge frequency at constant pressure in the chamber p = 16 Pa. (1) Experiment [4], (2) calculation [4], and (3) this work.

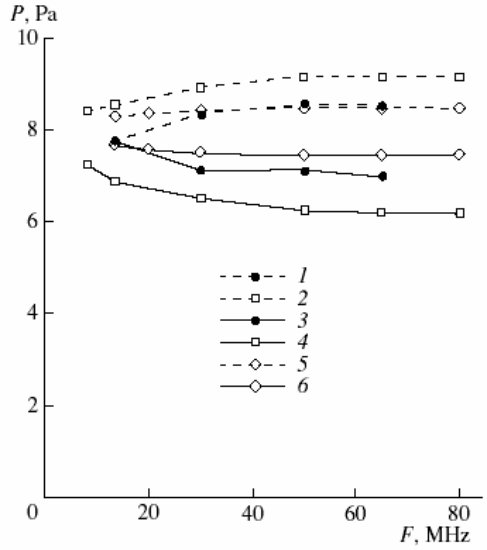


Fig. 2.4. Partial pressures of molecular hydrogen (dashed lines 1, 2, 5) and silane (solid lines 3, 4, 6) as a function of discharge frequency. Filled circles (1, 3): experiment [4]. Squares (2, 4): calculation [4]. Diamonds (5, 6): this work.

2.7.2. Specific features of chemical kinetics at low pressures

With the model described in this chapter, calculations of the processes are possible at low pressures, at which silane decomposition becomes noticeable and the concentration of molecular hydrogen can no longer be considered small. Below, we discuss in detail the qualitative differences of these low-pressure processes from those taking place at high pressures.

Dependence of the electron density in the discharge zone on the process parameters was described by a formula derived for the case of constant discharge power [3]

$$n_e\left(p_{SiH_4} l\right) = n_e^0 p_{SiH_4}^0 l^0 / p_{SiH_4} l\,, \tag{2.25}$$

where n_e^0 is the electron density under basic conditions, taken equal to $5 \cdot 10^8$ cm^{-3}.

This formula is true for pressures that are not very low, but it was used for the qualitative analysis in the entire range of pressures. For other key parameters of the problem, values characteristic of real experimental setups were chosen [1,2]: the temperature in the reaction chamber was taken equal to 520 K, the pressure was 0.25 Torr, and the inter-electrode distance was $l = L$ = 2.5 cm. Under these conditions, the reactor parameters corresponded to a characteristic pumping time of τ = 1s at a pressure of 1 Torr (for more details see Section 2.7.4). The calculations were carried out for a pressure range $0.02 < p < 1$ Torr. All the results given in this section refer to the case of pure silane as the carrier.

One of the main distinctions of the considered processes is a considerable rise in the concentration of atomic hydrogen with decreasing pressure. There are a number of reasons for this. First of all, as seen from formula (2.18), the concentration of atomic hydrogen in the particular case of homogeneous discharge is given by formula (2.26), i.e., it depends only on the electron density:

$$n_H = 2k_2 n_e / k_8 . \quad (2.26)$$

This formula is also in good agreement with the calculation results. At the same time, as seen from formulas (2.13), (2.14), and (2.25), the electron density and the constant k_2 rise rapidly with decreasing silane pressure, so that the concentration of atomic hydrogen increases by nearly five orders of magnitude as the pressure in the chamber decreases from 1.0 to 0.02 Torr and the relative silane concentration decreases to 0.3 as a result of its decomposition. The increase in atomic hydrogen concentration causes qualitative changes in various processes.

The behavior of other components with variations in pressure does not change compared with the description in [3], except for the fact that the concentration of Si_2H_6 drops at pressures below 0.1 Torr. The same effect is observed for Si_3H_8. This is due to a noticeable drop in silane concentration and the increase in atomic hydrogen concentration, the main reactants that determine the formation and decomposition of these components. The increase in the atomic hydrogen concentration also affects the balance and role of different reactions in silyl decomposition. Thus, at low pressures, an increase in the role of the decomposition process in the silyl balance is observed; at 0.02 Torr, it becomes dominant and prevails over deposition. Note, however, that this result is entirely a property of the model, because the use of formula (2.25) at very low pressures gives values of the electron density that are too high.

At pressures below about 0.08 Torr in silyl decomposition, the reaction $SiH_4 + Si_3H_8 \rightarrow Si_4H_9 + H_2$ gives way as the dominant process to reaction $SiH_3 + SiH_3 \rightarrow SiH_4 + SiH_2$; and at pressures of about 0.02 Torr, to the reaction between silyl and atomic hydrogen, which begins to make a noticeable contribution to the production of silyl. As before, silyl formation is completely dominated by the reaction between silane and atomic hydrogen.

Contributions of individual reactions to the silylyl balance remain largely unchanged, but its production at low pressures increases so much that most of the silylyl is deposited instead of being decomposed. Because of all these changes, the contribution of silylyl to film growth, which, with decreasing pressure, increases at a greater rate than that of silyl, becomes dominant at a pressure of 0.02 Torr. This should cause appreciable modifications in the film structure.

Taking into account silane decomposition at low pressures revealed further details of the division [3] of all components into three groups: stationary, nonstationary, and quasi-stationary. It was found that SiH can be considered strictly stationary only at high pressures. In addition, at high pressures, stationary components also include silylyl, the balance of which is maintained by two reactions with silane: $SiH_4 + e \rightarrow SiH_2 + 2H + e$ and $SiH_4 + SiH_2 \rightarrow Si_2H_6^*$, with the latter reaction keeping the concentrations of $Si_2H_6^*$ and Si_2H_4 (which is produced via decomposition of $Si_2H_6^*$) stationary. At low pressures, these last two components go over to the quasi-stationary group because of the appreciable

decomposition of silane, resulting in an increase in electron density and reaction rate constants. This is especially noticeable in the behavior of atomic hydrogen, whose concentration at high pressure is stationary and adequately described by formula (2.26). At low pressures, the deviation from stationarity caused by these processes makes atomic hydrogen non-stationary and appreciable accumulation of this component in the reaction chamber begins. Other parts of this classification remain qualitatively unchanged (see details in [3]).

2.7.3. Influence of diffusion coefficients

The use of formula (2.9) in calculating the diffusion coefficients does not significantly change the main results (such as the growth rate and film composition) compared with calculations in which these coefficients were assumed to be equal to the coefficients of the respective components in silane, as was done in [3]. This result appears to be quite evident at high pressures, where the concentration of molecular hydrogen is rather low and the diffusion coefficients calculated by formula (2.9) are close to the corresponding ambipolar coefficients. However, at low pressures, these values are distinctly different.

This fact suggests that the concentration profiles of the depositing components adjust in such a way as to maintain the production–deposition balance, irrespective of the particular values of the diffusion coefficients (the rate of chemical decomposition of silyl and silylyl at low pressures is small).

For a more detailed investigation of this issue, calculations using the above approach were carried out for a mixture of silane and molecular hydrogen at equal partial pressures and constant electron density $5 \cdot 10^8$ cm^{-3}. In this case, the influence of the diffusion coefficients is more pronounced. For the same purpose, a simplified formula was used to calculate reaction rate constants as functions of pressure [3]:

$$\frac{k_r}{k_r^0} = \left(\frac{l^0 p^0}{l_p} \right)^{2\epsilon_r/\epsilon_i} . \qquad (2.27)$$

Under these conditions, the diffusion coefficients calculated by Wilke's formula differ from the ambipolar diffusion coefficients by a factor of 1.6–1.7. Still, at low pressures where the silyl–silylyl balance is dominated by deposition, the fluxes of these components to the substrate practically coincide in the two cases, despite the distinctly different concentration profiles (see Fig. 2.5). At high pressures, silylyl is spent mainly through reactions in the bulk and the values of fluxes obtained with the use of constant ambipolar diffusion coefficients are somewhat less than those given by Wilke's formula; however, this difference is a mere 10% for silyl and 20% for silylyl, which is much less than the difference in the coefficients, and the contribution of silylyl to film growth under these pressures is insignificant.

To check this, a series of calculations was carried out for varying values of the diffusion coefficients. In one of these, the diffusion coefficient for silyl was assumed to be twice as large. The film growth rate was almost unchanged, although at a pressure of 1 Torr, the silyl concentration dropped by a factor of 1.5.

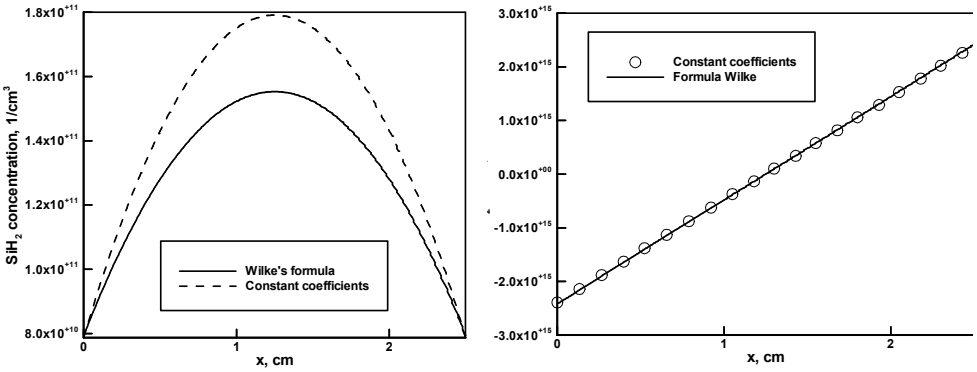


Fig. 2.5. SiH_2 concentration profiles (left) and SiH_2 fluxes (right) obtained with constant coefficients and using Wilke's formula. Pressure is 0.02 Torr

This fact is explained as follows. Because the dominant contribution to film growth comes from the flux of silyl, the growth rate can be analyzed using expression (2.20). At $d = 0$, this expression is significantly simplified, reducing to $\Gamma_{SiH_3} = 2k_2 n_e n_{SiH_4} L_3 \tanh(L/2L_3)$. For real values of the diffusion coefficient, $L/2L_3 < 0.7$ throughout the range of pressures studied and, therefore, the dependence of Γ_{SiH_3} on L_3 is weak. An increase in the diffusion coefficient only reduces the dependence of the flux of silyl on its value. Thus, silyl deposition is virtually unaffected by diffusion processes.

A similar result was also obtained for silylyl. The twofold increase in the silylyl diffusion coefficient did not produce any significant changes. At low pressures, the silylyl concentration decreased, but the modified profiles ensured exactly the same value of its flux. At higher pressures, the silylyl balance is completely controlled by chemical reactions; therefore, its concentration (as well as concentrations of all other components) remained the same, but the flux increased somewhat (by 37% at p = 1 Torr). However, at this pressure, the contribution from SiH_2 to film growth is quite small and cannot change the final result.

It can be concluded from the above analyses that further refining the diffusion transport description will not significantly change the results, especially in case of pure silane plasma.

2.7.4. Effect of the flow rate

The flow rate of the mixture through the working reactor volume is one of the important factors affecting the film growth parameters, and, at the same time, it can be easily varied. In [3], an approach that allows one to take into account the effect of the flow rate using a simple one-dimensional model was suggested. In this approach, the initial steady-state problem is replaced with a transient problem solved for a time period τ equal to the characteristic time of transit of the mixture in the reactor working volume. Decomposition of a considerable part of silane occurring at low flow rates and a significant increase in the molecular hydrogen concentration render the model suggested in [3] inapplicable for studying the role of the flow rate in a relatively wide range of this parameter.

Using the model presented in this study, the effect of the flow rate has been analyzed for the example of pure silane for τ ranging from 0.13 to 2.6 s, which, at a pressure of 0.25 Torr, corresponds to a change in the chamber volume from 1 to 20 l for the fixed flow rate of silane of 5 l/h at atmospheric pressure. Values of the other parameters were the same as in the previous sections, and the electron density was calculated by formula (2.25).

As in [3], variation of the flow rate was found to affect, first of all, the concentration of Si_2H_6: it increased by an order of magnitude as the flow rate was reduced by the same factor. At the same time, the concentration of Si_3H_8 increased by a factor of about 5. In the considered range of τ, the variation of the flow rate causes relatively small variation of the film characteristics. With a decreasing flow rate, the film growth rate increased by 30%. This was due to an increase in both the electron density and the reaction rate constants with a decreasing partial concentration of silane. The relative contribution of silylyl increased by about the same amount. This was caused by an increase in the concentration of molecular hydrogen and, subsequently, a greater contribution to the recovery of the silylyl concentration from the cyclic reaction $SiH_4 + SiH_2 \rightarrow Si_2H_6^* \rightarrow Si_2H_4 + H_2 \rightarrow SiH_4 + SiH_2$.

2.7.5. Effect of chamber volume

The problem considered above is closely related to the effect of the ratio of the reaction volume to the entire chamber volume, because circulation of the components throughout the volume can change the course of the processes.

To investigate this aspect in a one-dimensional case, the following approach was suggested: the area under study was expanded by adding a region R such that the ratio R_v would have been equal to the total chamber volume. The calculations for a fixed parameter τ and variable R_v gave estimates of the effect of circulation. The results described below were obtained for a total chamber volume of 2 l and $\tau = 0.26$ s; other parameters were the same as before.

Changing the parameter R_v from 1 to 0.2 resulted in an insignificant (about 25%) increase in the rates of film growth and silyl deposition. This increase was caused by increased concentrations of silane and silyl in the reaction volume because of the influx of SiH4 from the chamber volume. At the same time, silylyl deposition remained nearly constant, because some increase in its production due to the reaction of silylyl decomposition (R15) was offset by reaction $SiH_4 + SiH_2$ and deposition occurs only from a thin near-wall layer, where the concentration of silyl is low. Some distinctive features of these processes are illustrated in Fig. 2.6-Fig. 2.8.

In Fig. 2.6, profiles of typical representatives of the three classes of components are shown: non-depositing and slowly reacting (H_2), non-depositing and rapidly reacting (H), and depositing and rapidly reacting (SiH_2). It is seen that the concentrations of the latter two components in the reaction volume are practically independent of the chamber volume. They are determined by reactions proceeding in the discharge zone, and circulation has little effect on them. The concentration of molecular hydrogen, on the contrary, is constant throughout the chamber volume and is appreciably reduced by circulation.

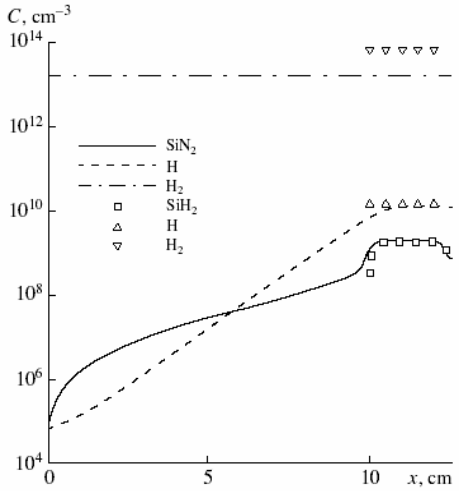


Fig. 2.6. Concentration profiles of silylyl and molecular and atomic hydrogen at different values of the ratio of reaction volume to the total chamber volume Rv equal to 0.2 (curves) and 1 (dots)

The small diffusion-reaction length for silylyl with respect to the chamber dimensions is the reason for the weak influence of the chamber volume on the balance and flows of this radical (Fig. 2.7, Fig. 2.8). Only some increase in its production due to a higher silane concentration can be noted; this increase is compensated by its more intensive (for the same reason) decomposition, so that flows onto the substrate turn out to be about the same.

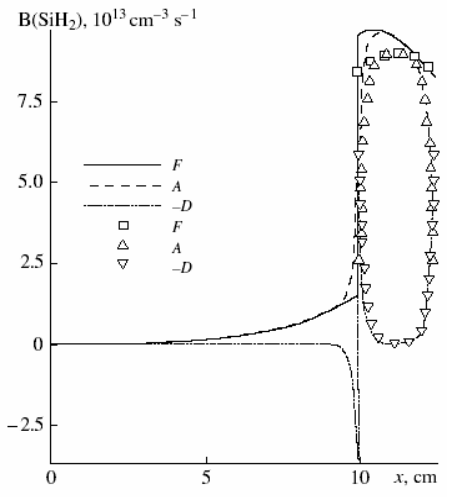


Fig. 2.7. Silylyl balance at different values of the ratio of reaction volume to the total chamber volume R_v. F, production; A, decomposition; D, diffusion; R_v = 0.2 (curves) and 1 (dots).

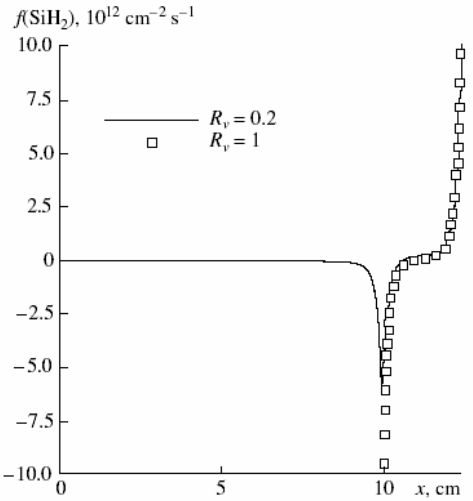


Fig. 2.8. Silylyl fluxes at different values of the ratio of the reaction volume to the total chamber volume R_v.

2.7.6. Effect of diluting silane with molecular hydrogen

In real experiments, a mixture of silane and molecular hydrogen is widely used. To evaluate the effect of silane dilution, calculations have been carried out for a silane content in this mixture varying from 100 to 20%, the other parameters having the same values as before. The electron density was determined using an RF discharge model [5]. Calculations for a discharge power of 0.025 W/cm^2 have shown that, in these conditions, the average electron density rises nearly linearly from $4.3 \cdot 10^8$ to $6.7 \cdot 10^8$ cm^{-3} as the silane content in the mixture is reduced.

As the silane content is reduced at constant pressure, the diffusion coefficients of the components, with the exception of H_2, increase by a factor of 2–3. Variation of the concentrations of all reaction products under these conditions is shown in Fig. 2.9. The film growth rate rises, and the relative contribution of silylyl to the overall deposition becomes larger (Fig. 2.10). The growth rate rises, despite falling silane concentration, because of an increase in the total production of depositing components due to higher rate constants of reactions initiated by electron impact (see expressions (2.13) and (2.14)) and increased electron density. In addition, the net result of lowering the concentration and the simultaneous increase in the diffusion coefficients is that the greater part of the produced silyl is deposited; this is true for other depositing components, except Si_2H_3 and Si_2H_4, which are effectively decomposed by molecular hydrogen. The behavior of silylyl is explained by the fact that with an increasing concentration of H_2 and a decreasing concentration of silane, the fraction of silylyl being deposited rises significantly. At the same time, the intensity of the above-mentioned cyclic reaction rises, leading to the recovery of the SiH_2 concentration.

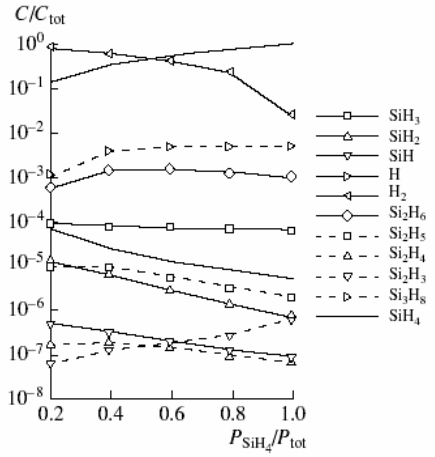


Fig. 2.9. Concentrations of components versus the silane fraction in the initial silane–hydrogen mixture at constant pressure.

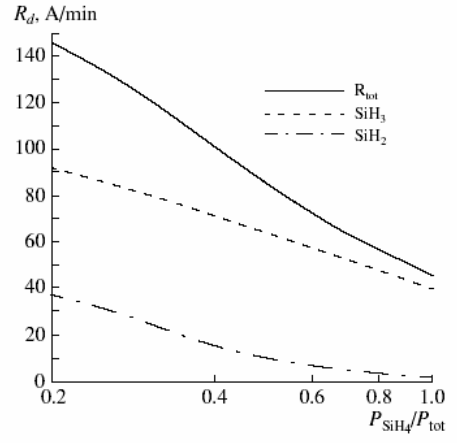


Fig. 2.10. Film growth rate and contributions of silyl and silylyl to it versus the fraction of silane in the initial silane–hydrogen mixture at constant pressure.

Also worth noting is the considerable reduction in the Si_3H_8 concentration with dilution (Fig. 2.9). The reason is that alongside the drop in production by the reaction $SiH_4 + Si_2H_4$, it is decomposed via the reaction with atomic hydrogen, whose concentration rises at a high rate. The result is a drastic drop in Si_4H_9 production, and the greater part of the silicon produced in the reaction is deposited.

2.7.7. Specific features of a triode system

All the results described above were obtained for a diode system in which the discharge zone is located between two electrodes, one of which serves as a substrate for the growing film (schematically shown in Fig. 1.1). One alternative technology is a so-called triode system, in which the discharge zone is separated from the substrate (see Section 2.1).

In order to investigate the effect of the separation of electrodes from the substrate, calculations of film growth from a pure silane plasma were carried out for the same problem parameters as previously used (with the exception of parameter L = 5 cm, with parameter l/L varying from 0.4 to 1). The average electron densities calculated with the RF discharge model [5] were varied from $6.8 \cdot 10^8$ to $4.9 \cdot 10^8$ cm^{-3}. The simulation results obtained using the above analytical expressions for the silyl flux (2.20) and the concentration profile of atomic hydrogen (2.18), (2.19) are in good agreement with numerical simulation results, as seen in Fig. 2.11 and Fig. 2.12, respectively.

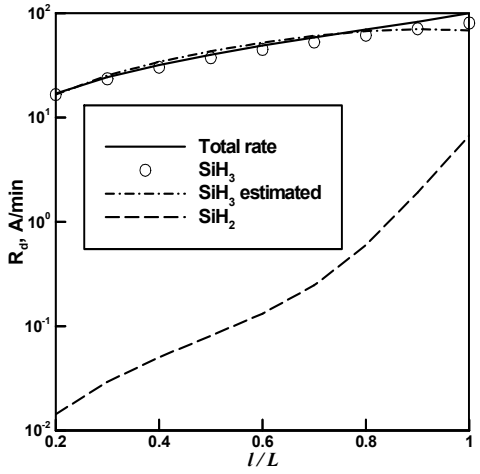


Fig. 2.11. Film growth rate and contributions of silyl and silylyl versus the ratio of discharge zone to the total reactor zone. Numerical and analytical solutions (formula (2.20)).

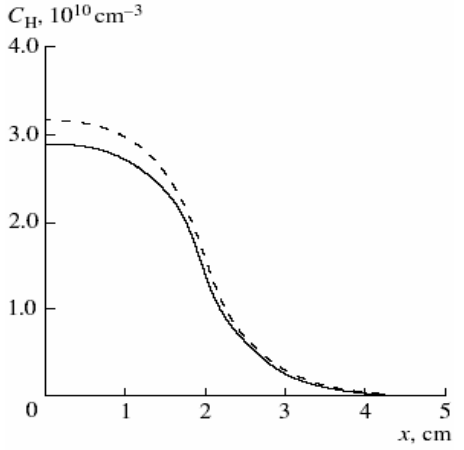


Fig. 2.12 Atomic hydrogen profiles for the triode reactor system (l/L = 0.4): solid curve, analytical solution; dashed curve, numerical solution.

As a result of varying the parameter l/L in the range specified above, the film growth rate (at a constant discharge power) decreased by only half. The cause of the slow decrease in the growth rate with increasing separation between the electrode and the substrate is the increase in electron densities and rate constants of reactions initiated by electron impact as the inter-electrode separation l is reduced, which partially compensates decomposition of

silicon-containing components outside the discharge zone. In addition, a drastic reduction in the contribution to the film growth from silylyl should be noted, starting as soon as the electrode is moved away from the substrate (Fig. 2.11).

This is especially remarkable in view of the fact that at the same time, its concentration and total production are rising. Two main reasons for this effect can be indicated. The first is that the diffusion-reaction length of silylyl is much less compared with that of silyl; therefore, the bulk decomposition of silylyl is more efficient. The second reason is their very different production mechanisms. The major source of silyl is the reaction between silane and atomic hydrogen, which, as seen in Fig. 2.12, diffuses quite intensively beyond the discharge zone.

On the other hand, the main contribution to silylyl production comes from the reaction between silane molecules themselves; therefore, production of SiH_2 outside the discharge is small.

One should pay attention to the rapid diminishing of high silane production that leads to the deposition of greater part of produced silicon (see Fig. 2.13).

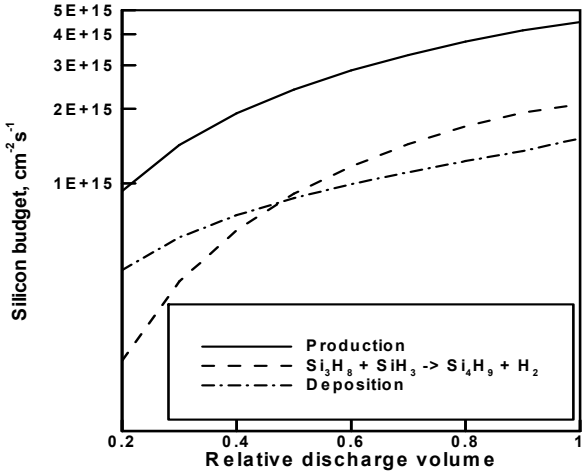


Fig. 2.13. Silicon budget.

2.7.8. Analysis of the diffusion-reaction lengths

Analysis of the flow in typical PECVD reactors showed that diffusion is the prevalent transport mechanism in the discharge volume between the parallel plate electrodes. To analyze the different behaviour of the particles, a diffusion-reaction length can be introduced [3] that characterizes the distance a particle Si_mH_n will diffuse over before undergoing chemical reaction:

$$L_{m,n} = (D_{m,n}/\nu_{m,n})^{1/2} \quad (2.28)$$

Here $D_{m,n}$ is the effective diffusion coefficient and $\nu_{m,n}$ is the removal rate of the *i*th species due to chemical reactions. Analysing diffusion-reaction lengths for different components of the mixture, one can get an idea of their spatial distribution; clarify their effect on the

composition of the mixture and separate out the boundary region, which gives a major contribution to the film growth. The value of this diffusion-reaction length determines the shape of the concentration profile between the electrodes: for low values the concentrations take the equilibrium values at all points in the chamber, falling abruptly to zero at the wall if the particles stick to the surface, or stay at equilibrium if they do not deposit. At high values of diffusion-reaction length the concentration distribution in space is smoothed out and weakly depends on local chemical reactions.

Fig. 2.14 shows the pressure dependence of diffusion-reaction lengths in the range of 0.15–1.00 Torr near two opposite electrodes (walls) maintained at temperatures T_w = (a) 300 and (b) 520 K, respectively [19], for the set of reactions listed in Table 2-3.

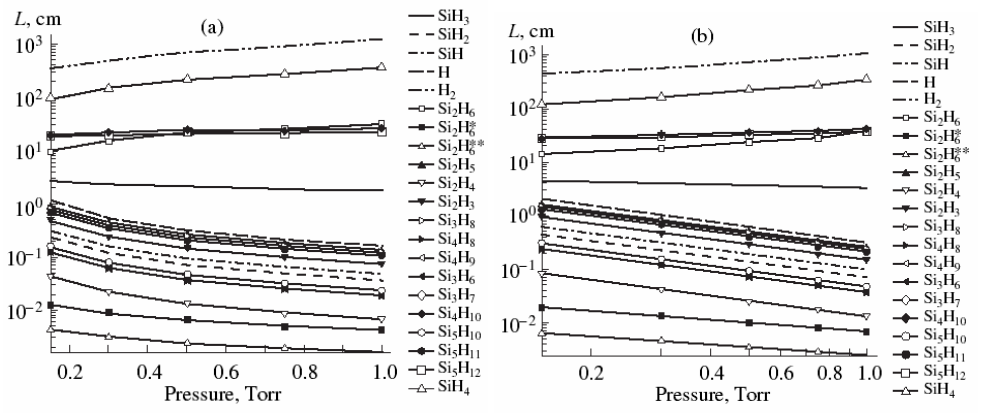


Fig. 2.14 Diffusion-reaction lengths vs. pressure at the (a) hot (Tw = 520 K) and (b) cold (Tw = 300 K) walls.

As we can see, all the components fall into three groups according to their diffusion-reaction lengths and the pressure dependences of these lengths. Molecules (particles with saturated bonds) fall into the first group. This group features relatively large diffusion-reaction lengths and their weak increase with pressure. For the species from this group, the reactions governing the removal rate $\nu_{m,n}$ are reactions with Si_2H_4 (molecular hydrogen); reactions with atomic hydrogen (and Si_2H_6); reaction with SiH_3 and at lower pressures with SiH_2 (species like Si_nH_{2n+2} with $n > 2$). Silyl (SiH_3) falls into the second group, because its diffusion-reaction length, first, differs from the corresponding values for the rest of the components and, second, slightly decreases with increasing pressure. This stems from the fact that silyl leaves the system not only through the reaction with itself but also through the reaction with silane. Since the effect of the latter reaction increases with pressure (when the silane concentration evidently rises), the removal rate turns out to be a function decreasing more slowly than $1/p$ (this is true for the pressure range considered; at higher pressures, it starts to increase). Accordingly, $L_{1,3}$ drops. The rest of the radicals fall into the third group. They feature relatively short diffusion-reaction lengths that decrease with increasing pressure. This is because the removal rate of these species is proportional to the concentrations of SiH_4 or H_2, which in turn are proportional to the pressure in the case at

hand and considerably exceed the concentrations of the other species. Thus, the corresponding lengths become inversely proportional to pressure.

The results of the film growth process were analyzed in [19]. A short summary of the key conclusions is presented here. Fig. 2.15 shows that the contribution from Si_2H_5 to the film growth is of the same order of magnitude as that from SiH_2. Our interest in the amount of the silylene deposit is explained by the fact that this radical is frequently viewed as the main cause for film degradation [3]. However, the other species are deposited in comparable (to silylene) amounts, making us attribute to them the same adverse effect. Analysis of the chemical reactions in the reactor suggests that a high effect of Si_2H_5 is associated with its relatively high diffusion-reaction length, which is approximately four times higher than that of SiH_2. In other words, the area of Si_2H_5 deposition is larger than that of silylene. The contribution from Si_3H_7 is appreciable; it is only 2.5 times lower than the contribution from Si_2H_5. A high efficiency of Si_3H_7 is related to its high concentration: the reaction of the Si_2H_4 radical with silane yields a high concentration of Si_3H_8 (R40: $Si_2H_4 + SiH_4 \rightarrow Si_3H_8$), which in turn produces Si_3H_7 in the reaction with silyl (R45: $Si_3H_8 + SiH_3 \rightarrow Si_3H_7 + SiH_4$). Thus, the noticeable contribution of Si_3H_7 to the film growth is explained by two factors: a high concentration of silane and a high reactivity of the Si_2H_4 radical. The diffusion-reaction length of Si_3H_7, as well as of Si_2H_5, is four times as large as that of SiH_2 and, accordingly, causes the same consequences.

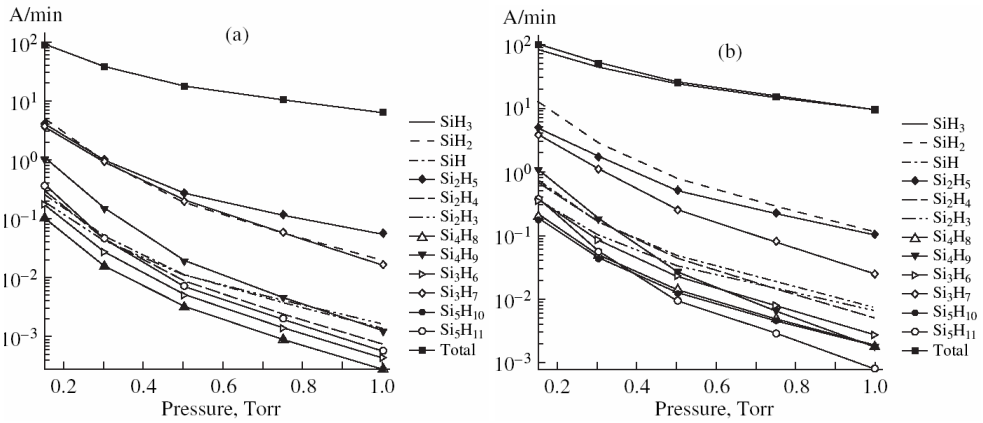


Fig. 2.15 Deposition rates of different components vs. pressure at the (a) hot (Tw = 520 K) and (b) cold (Tw = 300 K) walls.

Table 2-3. List of chemical reactions and reaction constants

Reaction no.	Reaction	K (cm^3/s)	Reaction no.	Reaction	K (cm^3/s)
R1	$SiH_4 + e \longrightarrow SiH_3 + H + e$	3.00(–11)[a, b]	R39	$Si_2H_4 + H_2 \longrightarrow Si_2H_6$	5.33(–13)
R2	$SiH_4 + e \longrightarrow SiH_2 + 2H + e$	1.50(–10)[b]	R40	$Si_2H_4 + SiH_4 \longrightarrow Si_3H_6 + H_2$	1.00(–11)[f]
R3	$SiH_4 + e \longrightarrow SiH + H + H_2 + e$	9.34(–12)[b]	R41	$Si_2H_4 + SiH_4 \longrightarrow Si_3H_8$	1.00(–10)
R4	$SiH_4 + e \longrightarrow SiH_2 + H_2 + e$	7.19(–12)[b]	R42	$Si_2H_3 + H_2 \longrightarrow Si_2H_5$	1.70(–12)
R5	$H_2 + e \longrightarrow 2H + e$	4.49(–12)[b]	R43	$Si_3H_8 + H \longrightarrow Si_2H_5 + SiH_4$	1.11(–12)[g]
R6	$Si_2H_6 + e \longrightarrow SiH_3 + SiH_2 + H_2 + e$	3.72(–10)[b]	R44	$Si_3H_8 + H \longrightarrow Si_2H_6 + SiH_3$	1.11(–12)[c]
R7	$Si_2H_6 + e \longrightarrow Si_2H_4 + 2H + e$	3.70(–11)[b]	R45	$Si_3H_8 + H \longrightarrow Si_3H_7 + H_2$	2.16(–12)[c]
R8	$SiH_4 + H \longrightarrow SiH_3 + H_2$	2.67(–12)[c]	R46	$Si_3H_8 + SiH_3 \longrightarrow Si_3H_7 + SiH_4$	2.49(–12)[c]
R9	$SiH_4 + SiH_3 \longrightarrow Si_2H_5 + H_2$	1.78(–15)	R47	$Si_3H_8 + SiH_2 \longrightarrow Si_4H_{10}$	1.20(–10)[c]
R10	$SiH_4 + SiH_2 \longrightarrow Si_2H_6^*$	1.00(–11)	R48	$Si_3H_8 + SiH \longrightarrow Si_4H_9$	1.00(–11)
R11	$SiH_4 + SiH \longrightarrow Si_2H_3 + H_2$	1.70(–12)	R49	$Si_3H_8 + Si_2H_5 \longrightarrow Si_3H_7 + Si_2H_6$	2.49(–12)[c]
R12	$SiH_4 + SiH \longrightarrow Si_2H_5$	2.50(–12)	R50	$Si_3H_8 + Si_2H_4 \longrightarrow Si_5H_{12}$	1.00(–10)[d]
R13	$SiH_3 + H \longrightarrow SiH_2 + H_2$	1.00(–10)	R51	$Si_3H_7 + H \longrightarrow Si_3H_8$	1.00(–11)
R14	$SiH_3 + SiH_3 \longrightarrow SiH_4 + SiH_2$	1.50(–10)	R52	$Si_3H_7 + SiH_4 \longrightarrow Si_3H_8 + SiH_3$	5.00(–13)[h]
R15	$SiH_3 + SiH_3 \longrightarrow Si_2$	1.00(–11)	R53	$Si_3H_7 + SiH_3 \longrightarrow Si_3H_6 + SiH_4$	1.00(–10)[i]
R16	$SiH_3 + SiH_2 \longrightarrow Si_2H_5$	3.77(–13)	R54	$Si_3H_7 + SiH_3 \longrightarrow Si_4H_{10}$	5.50(–11)[j]
R17	$SiH_2 + H \longrightarrow SiH + H_2$	7.96(–13)	R55	$Si_3H_7 + SiH_2 \longrightarrow Si_4H_9$	1.00(–11)
R18	$SiH_2 + H \longrightarrow SiH_3$	1.11(–12)	R56	$Si_3H_6 + SiH_4 \longrightarrow Si_4H_8 + H_2$	1.00(–11)[f]
R19	$SiH_2 + H_2 \longrightarrow SiH_4$	2.00(–13)	R57	$Si_3H_6 + SiH_4 \longrightarrow Si_3H_8$	1.00(–11)[k]
R20	$SiH_2 + SiH \longrightarrow Si_2H_3$	7.22(–13)	R58	$Si_3H_6 + SiH_3 \longrightarrow Si_4H_9$	2.00(–11)
R21	$SiH + H_2 \longrightarrow SiH_3$	1.98(–12)	R60	$Si_4H_{10} + SiH_3 \longrightarrow Si_4H_9 + SiH_4$	2.49(–12)[m]
R22	$Si_2H_6 + H \longrightarrow SiH_4 + SiH_3$	1.11(–12)[c]	R61	$Si_4H_{10} + SiH_2 \longrightarrow Si_5H_{12}$	1.20(–10)[n]
R23	$Si_2H_6 + H \longrightarrow Si_2H_5 + H_2$	1.20(–10)[c]	R62	$Si_4H_{10} + SiH \longrightarrow Si_5H_{11}$	1.00(–11)
R24	$Si_2H_6 + SiH_3 \longrightarrow Si_2H_5 + SiH_4$	1.00(–12)[c]	R63	$Si_4H_9 + H \longrightarrow Si_4H_{10}$	1.00(–11)
R25	$Si_2H_6 + SiH_2 \longrightarrow Si_3H_8$	1.20(–10)	R64	$Si_4H_9 + SiH_4 \longrightarrow Si_4H_{10} + SiH_3$	5.00(–13)[h]
R26	$Si_2H_6 + SiH \longrightarrow Si_3H_7$	1.00(–11)	R65	$Si_4H_9 + SiH_3 \longrightarrow Si_4H_8 + SiH_4$	1.00(–10)[i]
R27	$Si_2H_6 + Si_2H_4 \longrightarrow Si_4H_{10}$	1.00(–10)[d]	R66	$Si_4H_9 + SiH_3 \longrightarrow Si_5H_{12}$	5.50(–11)[j]
R28	$Si_2H_6^* + M^e \longrightarrow Si_2H_6 + M$	1.00(–10)	R67	$Si_4H_9 + SiH_2 \longrightarrow Si_5H_{11}$	1.20(–10)[n]
R29	$Si_2H_6^* \longrightarrow Si_2H_4 + H_2$	5.00(6) s^{-1}	R68	$Si_4H_8 + SiH_4 \longrightarrow Si_5H_{10} + H_2$	1.00(–11)[f]
R30	$Si_2H_6^{**} \longrightarrow SiH_4 + SiH_2$	2.30(7) s^{-1}	R69	$Si_4H_8 + SiH_4 \longrightarrow Si_5H_2$	1.00(–11)[k]
R31	$Si_2H_6^{**} \longrightarrow Si_2H_4 + H_2$	2.30(7) s^{-1}	R70	$Si_4H_8 + SiH_3 \longrightarrow Si_5H_{11}$	2.00(–11)
R32	$Si_2H_6^{**} + M^e \longrightarrow Si_2H_6 + M$	2.00(–10)	R71	$Si_5H_{12} + H \longrightarrow Si_5H_{11} + H_2$	2.16(–12)[l]
R33	$Si_2H_5 + H \longrightarrow Si_2H_4 + H_2$	1.00(–10)	R72	$Si_5H_{12} + SiH_3 \longrightarrow Si_5H_{11} + SiH_4$	2.49(–12)[m]
R34	$Si_2H_5 + SiH_4 \longrightarrow Si_2H_6 + SiH_3$	5.00(–13)	R73	$Si_5H_{12} + SiH_2 \longrightarrow Si_6H_{14}$	1.20(–10)[n]
R35	$Si_2H_5 + SiH_3 \longrightarrow Si_2H_4 + SiH_4$	1.00(–10)	R74	$Si_5H_{11} + H \longrightarrow Si_5H_{12}$	1.00(–11)
R36	$Si_2H_5 + SiH_3 \longrightarrow Si_3H_8$	5.50(–11)[c]	R75	$Si_5H_{11} + SiH_4 \longrightarrow Si_5H_{12} + SiH_3$	5.00(–13)[h]
R37	$Si_2H_5 + Si_2H_5 \longrightarrow Si_4H_{10}$	1.00(–11)	R76	$Si_5H_{11} + SiH_3 \longrightarrow Si_5H_{10} + SiH_4$	1.00(–10)[i]
R38	$Si_2H_4 + H_2 \longrightarrow SiH_4 + SiH_2$	1.00(–10)	R77	$Si_5H_{10} + SiH_4 \longrightarrow Si_6H_{14}$	1.00(–11)[k]

Notes: *a* 3.00(–11) = 3.00 . 10–11.

b Data for electron-impact reaction rates are given for the same reference conditions as in Table 2-2.

c The values of the constants are taken from [13].

d The values of the constants are taken by analogy with reaction R41 from [16].

e *M* is partial particle (collision partner).

f Reactions are proposed in [15].

The values of the constants were taken by analogy with:

g Reactions R22 and R44 in [13].

h Reaction R34 in [3].

i Reaction R35 in [16].

j Reaction R36 in [13].

k Reaction $SiH_2 + SiH_4$ Si_2H_6 taken from [13].

l Reaction R45 in [13].
m Reaction R46 in [13].
n Reaction R47 in [13].

2.8. Conclusions

In this work, numerical investigations of the growth of hydrogenated amorphous silicon films have been carried out under various conditions in the growth chamber.

Comparison with experimental data has demonstrated the ability of the model to predict the film growth rate and concentrations of individual components with reasonable accuracy.

It has been shown that the widely used technique of diluting silane with molecular hydrogen both increases the growth rate and reduces production of higher silanes, making this technology more economical and ecologically clean. Moreover, dilution increases the contribution of silylyl to film growth, appreciably affecting its properties.

Analysis of a widely used reactor system has been carried out. Numerical simulation has shown that the effective decomposition of silylyl outside the discharge zone reduces its contribution to film growth as the substrate is displaced away from the discharge.

The analytical expressions obtained for silyl and silylyl fluxes and for the profile of atomic hydrogen closely approximate the results of numerical computations and can be used for making the corresponding estimations.

In the next chapter we investigate the influence of flow on the deposition process in 2D and 3D geometries of a reactor chamber.

2.9. References

1. O. A. Golikova, Sov. Phys. Semicond. 25, 915 (1991).
2. M. J. Kushner, J. Appl. Phys. 63, 2532 (1988).
3. Yu. E. Gorbachev, M. A. Zatevakhin, and I. D. Kaganovich. Tech. Phys. 41, 1247 (1996)
4. G. J. Nienhuis, W. J. Goedheer, et al., J. Appl. Phys. 82, 2060 (1997).
5. V. A. Schweigert, M. I. Zhilyaev, I. V. Schweigert, Applied Mechanics and Technical Physics. 35 (1), 13 (1994).
6. C. R. Wilke, Chem. Eng. Prog. 46 (2), 95 (1950).
7. J. O. Hirschfelder, C. F. Curtiss, and R. B. Bird, Molecular Theory of Gases and Liquids (Wiley, New York, 1954; Inostrannaya Literatura, Moscow, 1961).
8. P. Ginzburg, Friction and Heat Transfer in Moving Gas Mixtures (Leningrad Univ., Leningrad, 1975).
9. J. Perrin and O. Leroy, Contrib. Plasma Phys. 36 (3), 3 (1996).
10. N. Itabashi, K. Kamo, and N. Nishawaki, Jpn. J. Appl. Phys. 28, 325 (1989).
11. J. Perrin, J. Phys. D 26, 1662 (1993).

12. Yu.E. Gorbachev, M.A. Zatevakhin, V.V. Krzhizhanovskaya, V. A. Shveigert. Special Features of the Growth of Hydrogenated Amorphous Silicon in PECVD Reactors. Technical Physics, Vol. 45, N 8, pp. 1032-1041. Publ.: MAIK Nauka/Interperiodika 2000
13. Yu.E. Gorbachev, M.A. Zatevakhin, V.V. Krzhizhanovskaya. Influence of Silane Dilution with Hydrogen and Inert Gases on the Dynamics of Amorphous Hydrogenated Silicon Films Growth in PECVD Reactors. Proceedings of the 2nd International Conference on Amorphous and Microcrystalline Semiconductors, St. Petersburg, Russia, 2000
14. Y.E. Gorbachev, M.A. Zatevakhin, V.V. Krzhizhanovskaya. Hydrogenated Amorphous Silicon Film Growth in Remote PECVD. Proceedings of the All-Russian Symposium on Amorphous and Microcrystalline Semiconductors, St. Petersburg, Russia, 5-9 July 1998, p.24
15. Yu.E. Gorbachev, M.A. Zatevakhin, V.V. Krzhizhanovskaya. Simulation of Amorphous Hydrogenated Silicon Plasma Enhanced Chemical Vapor Deposition in a Triode Reactor. Proceedings of the NATO workshop on Plasma Physics, St. Petersburg, Russia, 1997
16. Nienhuis, G.J., 1998, *Plasma models for silicon deposition*. PhD Thesis, Utrecht University.
17. Leroy, O., Gousset, G., Alves, L.L., Perrin, J., Jolly, J., 1998, Plasma Sources Sci. Technol., **7**, p. 348
18. K. De Bleecker, A. Bogaerts, R. Gijbels, W. Goedheer. Numerical investigation of particle formation mechanisms in silane discharges. PHYSICAL REVIEW E 69, 056409 (2004)
19. Yu.E. Gorbachev. Effect of Oligomers on the Growth of Amorphous Silicon Films in a PECVD Reactor. Technical Physics, 2006, Vol. 51, No. 6, pp. 733–739.
20. V. V. Krzhizhanovskaya, P.M.A. Sloot and Yu. E. Gorbachev. Grid-based Simulation of Industrial Thin-Film Production. Simulation: Transactions of the Society for Modeling and Simulation International, Special Issue on Applications of Parallel and Distributed Simulation in Industry, January 2005, V. 81, No. 1, pp. 77 - 85

Chapter 3. Modeling and Simulation Part II: 2D and 3D Flow and 1D Plasma Discharge*

3.1. Introduction

In this chapter we follow the same modeling approach for coupling the models as in previous chapters: The plasma model is applied in the discharge zone and the reactive flow model is used in the entire reactor volume (see Section 2.1). To study the influence of the reactor configuration on the flow field and on the deposition homogeneity, we develop a 2D reactive flow model suitable for simulating axisymmetric reactors. For the asymmetric geometries, a 3D flow model is also developed as an extension of 2D. The purpose of this Chapter is to investigate only the influence of the flow dynamics, and not the spatial effects of plasma, so we keep the 1D plasma model as in the previous chapter.

3.2. 2D and 3D flow models

The model presented in this section is an extension of the one-dimensional model [9,10] previously developed by the authors. In the whole reactor domain we solved the two- or three-dimensional Navier-Stokes equations for the flow of a viscous compressible multi-component chemically active gas mixture. More details on the models are given below.

3.2.1. Governing equations for reactive flow

We use computational fluid dynamics to simulate the physical and chemical processes that occur in PECVD reactors. The governing Navier-Stokes equations for an unsteady laminar flow of viscous compressible multi-component mixture of chemically reacting gases can be written as [11]:

$$\frac{d}{dt}\int_V U dV + \oint_S (F+G)\cdot dS = \int_V H dV \,, \tag{3.1}$$

where t is time and $U = (\rho,\ \rho u_x\ ,\ \rho u_y\ ,\ \rho u_z\ ,\ \rho e,\ \rho f_1\ ,\ldots,\ \rho f_{N-1}\ )^T$ is a vector of conservative variables (here ρ is the gas density, u_x, u_y, u_z are the velocity vector components, $e = e_{\text{int}} + \left(u_x^2 + u_y^2 + u_z^2\right)/2$ is the total energy per mass unit and f_i is the mass fraction of the i^{th} species, $\sum_{i=1}^{N} f_i = 1$), V is the arbitrary gas volume enclosed by the surface S, and $d\mathbf{S}$ is the unit vector perpendicular to the surface element, pointing outward. The internal energy is given by $e_{\text{int}} = \sum_{i=1}^{N} f_i\, e_{\text{int}\,i}$, $e_{\text{int}\,i} = h_{0i} + \int_{T_0}^{T} C_{\text{p}i}\, dT - R_i T$. Here h_{0i} is the heat of

* Parts of this chapter were published in [1-8]

formation at T_0 =293 K, C_{pi} is the heat capacity at constant pressure p, and R_i is the gas constant of the i^{th} species.

The flux vector $\mathbf{F} = F_x\mathbf{i} + F_y\mathbf{j} + F_z\mathbf{k}$ is associated with the inviscid gas transfer, and $\mathbf{G} = G_x\mathbf{i} + G_y\mathbf{j} + G_z\mathbf{k}$ describes viscous transfer, energy dissipation, heat conductivity and diffusion:

$$F_x = \begin{Bmatrix} \rho u_x \\ \rho u_x^2 + p \\ \rho u_x u_y \\ \rho u_x u_z \\ (\rho e + p) u_x \\ \rho u_x f_1 \\ \dots \\ \rho u_x f_{N-1} \end{Bmatrix} \quad G_x = \begin{Bmatrix} 0 \\ -\sigma_{xx} \\ -\tau_{xy} \\ -\tau_{xz} \\ -u_x\sigma_{xx} - u_y\tau_{xy} - u_z\tau_{xz} + q_x \\ J_{1x} \\ \dots \\ J_{(N-1)x} \end{Bmatrix}$$

Similar expressions hold for F_y, F_z, G_y, G_z.

The viscous stress tensor elements are given by $\sigma_{xx} = 2\mu\left(\frac{\partial u_x}{\partial x} - \frac{1}{3} div\mathbf{u}\right)$, $\tau_{xy} = \mu\left(\frac{\partial u_x}{\partial y} + \frac{\partial u_y}{\partial x}\right)$. The heat flux vector is $\mathbf{q} = -\lambda \cdot grad(T) + \sum_{i=1}^{N} \mathbf{J_i} h_i$, where h_i is the enthalpy of the i^{th} species per mass unit and $\lambda = \mu C_p / Pr$ is the mixture heat conductivity, and Pr is the Prandtl number. The diffusive flux vector is given by $\mathbf{J_i} = -\rho D_i grad(f_i)$. The transport coefficients, dynamic viscosity μ and diffusion coefficients D_i, are calculated from the molecular kinetic theory of gases using the Lennard-Jones 6-12 potential [10,12].

The source term $H = (0,...,0,H_1,\dots,H_N)^T$ in eq. (1) describes the species production due to the chemical reactions, with $H_i = \frac{M_i}{N_A}[B_i - A_i n_i]$, $A_i = \sum_{j=1}^{N}\left(n_e k_{ij} + \sum_{l=1}^{N} K_{jli} n_l\right)$, $B_i = \sum_{j=1}^{N} n_j\left(n_e k_{ij} + \sum_{l=1}^{N} K_{jli} n_l\right)$, where k_{ij} and K_{jli} are reaction rate constants for electron-induced reactions and neutral-neutral reactions respectively, N_A is the Avogadro number, n_i is the number density and M_i is the molar mass of the i^{th} species.

3.2.2. Boundary conditions for the reactive flow model

Under the low gas densities involved in the reactor (of order of 10^{-3}-10^{-4} kg/m^3), the Knudsen number (ratio of the mean free path of the radicals to the inter-electrode distance) is of the order of 0.01, at which the continuum approach starts to fail. To take into account this effect, a slip velocity and temperature jump boundary conditions at the wall should be applied. The effect of the velocity slip is probably negligible under most of the conditions for which simulations were done. But the effect of thermal accommodation is definitely important, since the accommodation coefficients for the diluent hydrogen are known to be small, particularly for SiC [13,14]. It was shown that at low pressures (near or below 100

Pa), the gas can be substantially colder than the surface at a distance less than one mean free path from that surface. This drastically lowers the rate of chemical reactions in the reactor volume with respect to the case where accommodation would be complete [15].

In the local 2D coordinate system x'-y' with the axis y' perpendicular to the wall (see Fig. 3.1) the velocity slip boundary condition is given by [16]:

$$u'_x = \frac{2-\theta}{\theta}\frac{\partial u'_x}{\partial y'}\frac{\mu}{\rho}\sqrt{\frac{\pi}{2R_u T}}\Bigg/\sum_{i=1}^{N}\left(\frac{f_i}{M_i}\right),\ u'_y = 0.$$

Here R_u is the universal gas constant and θ is an accommodation coefficient (close to 1).

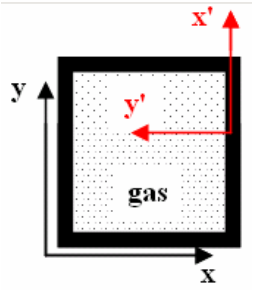


Fig. 3.1. Local coordinates

The temperature jump boundary condition at the wall is given by:

$$T = T_w + \frac{T_w}{T}\frac{\partial T}{\partial y'}\frac{1}{2R_u}\sqrt{\frac{\pi}{2R_u T_w}}\frac{2-\theta}{\theta}\lambda/\rho\sum_{i=1}^{N}\left(\frac{f_i}{M_i^{3/2}}\right),$$

where T_w is the wall temperature.

The boundary condition for the species concentration at the wall was derived in [9]: $\frac{\partial \rho f_i}{\partial y'} = \frac{s_i\sqrt{2R_i T/\pi}}{(2-s_i)D_i}\rho f_i$, where s_i = 0.0 for the molecules Si_nH_{2n+2}, s_i = 0.15 for the radicals with one free bond Si_nH_{2n+1} and s_i = 1 for the radicals with more than one free bond.

The deposition rate is calculated as follows [10]: $R_f = -N_A\, v\sum_{i=1}^{N}\frac{q_i D_i}{M_i}\frac{\partial \rho f_i}{\partial y'}$, where $v = 2.195\cdot 10^{-29}\ m^3$ is the silicon atom volume, and q_i is the number of silicon atoms in a depositing radical.

For the inlet boundary conditions, the flow rate, temperature and mixture composition need to be specified. The outlet pressure is set equal to the gas pressure inside the reactor chamber, and for all the other variables a "soft" boundary condition is used: $\partial/\partial y' = 0$.

3.3. Numerical methods and algorithms

3.3.1. Numerical scheme

To solve the set of equations (3.1) we used the operator splitting technique, decoupling the physical processes [17]. According to this method the problem can be split into inviscid and viscous "steps":

$$\frac{\partial}{\partial t}\int_V U^{**}dV + \oint_S G(U^{**})\cdot dS = \int_V H dV,\quad U^{**}(t) = U^{*}(t+\Delta t)$$

$$\frac{\partial}{\partial t}\int_V U^{*}dV + \oint_S F(U^{*})\cdot dS = 0,\quad U^{*}(t) = U(t)$$

The stiffness of the system is resolved by implicit numerical schemes. At the first step, an implicit ENO-scheme [18] is used in combination with a bi-diagonal algorithm [19]

for implicit increment calculation. For solving the "viscous" part, an implicit finite-difference scheme [20] is used. Given these finite-difference schemes we need to define the actual computational mesh.

3.3.2. Multi-block mesh generation

Since a PECVD reactor has a complex shape, a multi-block mesh generation algorithm is used, where the complete computational domain is divided into simple blocks, and in each block a regular mesh is generated. A mechanical analogy with deformed bodies [4] is applied for generation of non-uniform meshes in non-rectangular blocks, see Fig. 3.2. An irregular block connection was used for blocks with non-coinciding mesh vertices.

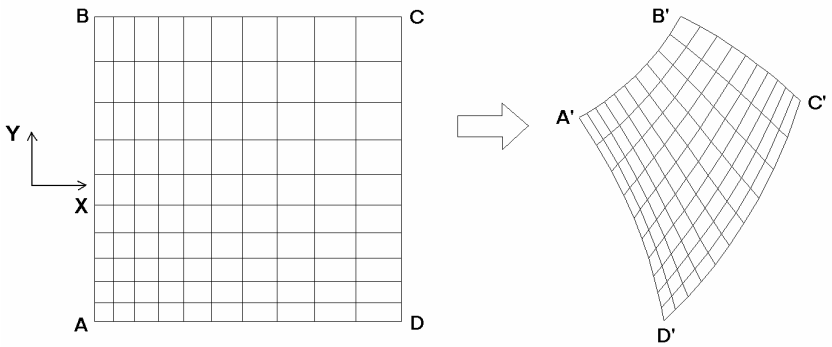


Fig. 3.2. A mechanical analogy with deformed bodies for generation of non-uniform meshes in non-rectangular blocks

3.3.3. Parallelization

Two levels of parallelization have been applied to solve the problem efficiently on computer clusters: job-level parallelization (parameter sweeps) and task-level parallelization (domain decomposition). The task-level parallel algorithm used beam distribution among the processors. A detailed description of parallelization methods will be given in Section 5.2.

3.4. Implementation and Problem Solving Environment components

For implementation of the numerical scheme and parallel algorithms, the MPI message passing interface was used in combination with a computational C++ core. The graphical user interface (GUI) was designed with the use of C/C++ programming languages, widespread platform independent GTK+ graphics library (glib, gdk, gtk) [21] and the Glade user interface builder [22]. The developed application provides users with an intuitive interface that allows them to interactively control the simulation processes, to visualize numerical results in real-time (or postponed) mode, and to access chemical components data or the archives of results. All modules of the Virtual Reactor problem

solving environment are platform-independent and were tested on Unix-like and Microsoft Windows operating systems.

To provide access to the Virtual Reactor, an interface for remote use was developed based upon Web technology, using a client-server model. For this interface, we created HTML pages, Java applets, JavaScript and CGI scripts. This system, like a local GUI, provides control over simulations and graphical representation of physical parameters, and was integrated into the Migrating Desktop.

To utilize the advantages of the parameter-sweep approach, a task Manager system was developed, looking after cluster processor idleness and loading them with new jobs. When one of the processors finished its job, the Manager initiates a process with new initial data set, running over the values of parameters studied. The jobs are submitted through the Migrating Desktop Grid-portal and the job management is done by the Resource Manager of the Grid environment. More details on the Virtual Reactor PSE are given in Chapter 6.

3.5. 2D Simulation Results

3.5.1. Comparison with experiment and investigating possible causes of mismatch

The simulations of 2D flow with 1D plasma (where the 1D electron distribution is simply extrapolated along the 2D electrodes) predict a "convex" shape of film thickness, with maximum in the center and minimum at the edges (see for instance Fig. 3.3). The experimentally observed film thickness distribution however, was "concave, with a minimum in the center and a maximum near the edges [23].

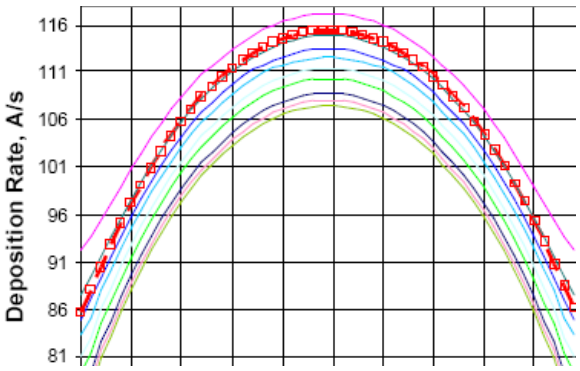


Fig. 3.3. Simulation results of deposition rate distribution along the substrate.

A number of test experiments were performed to find out which parameters can be responsible for this mismatch in the film distribution. A little non-uniformity in substrate temperature distribution (actually measured in experiments) does not influence the deposition rate too much: the dip in the middle is only of 1-2% (see Fig. 3.4, left). The electron density non-uniformity (artificially introduced in the simulation) does influence the deposition significantly (see Fig. 3.4, right). This is a clear indication that a 2D plasma model is needed to correctly predict the film distribution along the substrate.

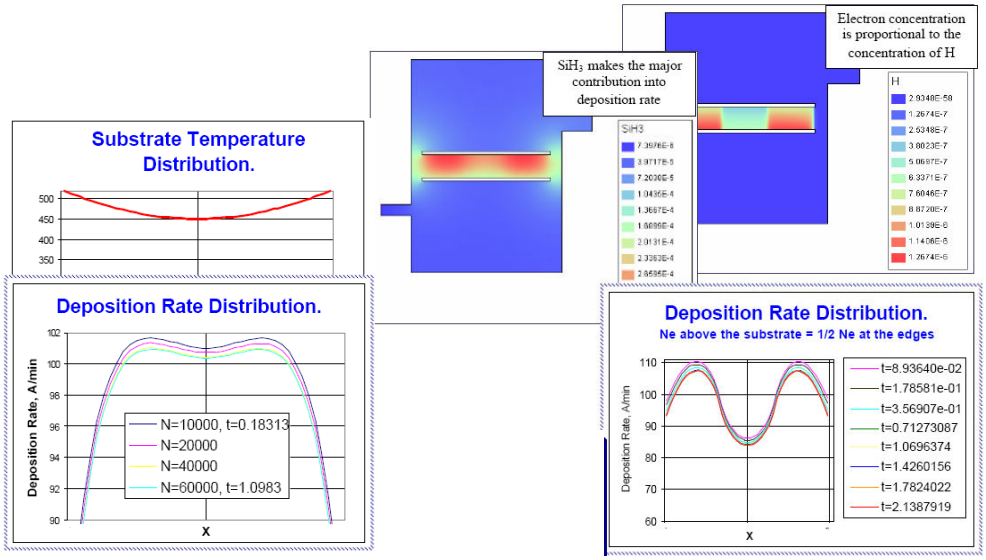


Fig. 3.4. Influence of non-uniform distribution of substrate temperature (left) and electron concentration (right) on the film distribution along the substrate.

3.5.2. Influence of the flow rate, temperature and discharge power

In the following experiments the inter-electrode spacing is 5 cm, gas temperature inside the chamber and at the inlet 300 K, substrate temperature 523 K, pressure 0.15 Torr (20 Pa), power supplied to the electrode 5 W and the discharge frequency 50 MHz. Equal molar fractions of SiH_4 and H_2 are injected, with a nominal total flow rate of 20 sccm [4]. The gas mixture enters the chamber at the left side of the simplified chamber geometry shown in Fig. 3.5.

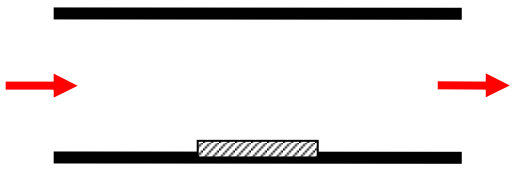


Fig. 3.5. Simplified design of a 2D reactor chamber.

The simulation showed that tripling the flow rate (from 20 sccm to 60 sccm) reduces the film growth rate, and the deposited film becomes essentially asymmetric, with a noticeably higher growth rate at the downwind side of the substrate (see Fig. 3.6). Another remarkable observation is that the contribution of higher silanes (Si_3H_7) to deposition decreases when flow rate increases. This is explained by the relatively low reactivity and high diffusion-reaction length of this species (four times as large as that of SiH_2), which

makes it travel over long distances without chemical transformations and be removed from the volume above the substrate. Decreasing the flow rate leads to a more uniform and symmetrical film profile and a higher contribution of Si_3H_7 into the resulting film.

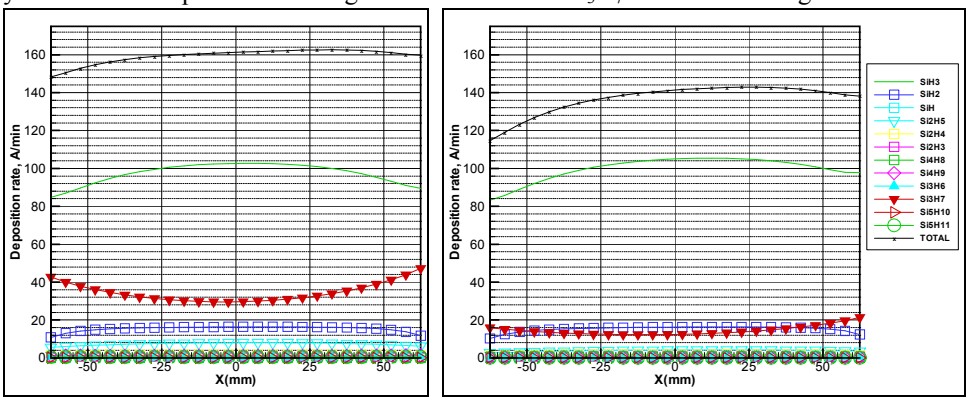


Fig. 3.6. Contribution of different species into deposition. Pressure p=0.15 Torr, discharge power W=0.05 W/cm2. Flow rate is 20 sccm (left) and 60 sccm (right).

Influence of pressure on the deposition rate is demonstrated in Fig. 3.7 for two different surface temperatures. Increasing the pressure significantly slows down the deposition process, keeping the relative contribution of different species to the deposition approximately constant. The same trend is observed in real experiments for pressures up to 1.5-2 Torr. This can be explained by the fact that pressure most strongly influences the plasma discharge and electron concentration, which in turn affects the radical production in electron-impact reactions R1-R7 and in reactions with atomic hydrogen (see Table 3-1).

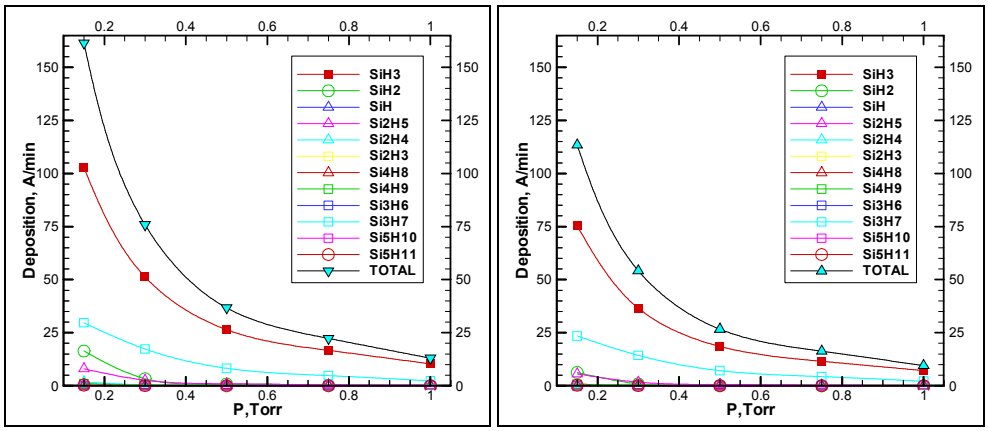


Fig. 3.7. Influence of pressure on the deposition rate. Discharge power W=0.05 W/cm^2.
Left: Deposition in the middle of the hot substrate (T=520 K).
Right: Deposition in the middle of the cold electrode (T=300 K).

Table 3-1. List of chemical reactions and reaction constants used in simulations in this Section. Data for electron-impact reaction rates are given for the same reference conditions as in Table 2-2.

	Reaction	Reference reaction rate, cm^3/s
R1	SiH4 + e → SiH3 + H + e	3.00e-11
R2	SiH4 + e → SiH2 + 2H + e	1.50e-10
R3	SiH4 + e → SiH + H + H2 + e	9.34e-12
R4	SiH4 + e → SiH2 + H2 + e	7.19e-12
R5	H2 + e → 2H + e	4.49e-12
R6	Si2H6 + e → SiH3 + SiH2 + H + e	3.72e-10
R7	Si2H6 + e → Si2H4 + 2H + e	3.70e-11
R8	SiH4 + H → SiH3 + H2	2.53e-12
R9	SiH4 + SiH3 → Si2H5 + H2	1.78e-15
R10	SiH4 + SiH2 → Si2H6*	1.00e-11
R11	SiH4 + SiH → Si2H3 + H2	1.70e-12
R12	SiH4 + SiH → Si2H5	2.50e-12
R13	SiH3 + H → SiH2 + H2	1.00e-10
R14	SiH3 + SiH3 → SiH4 + SiH2	1.50e-10
R15	SiH3 + SiH3 → Si2H6**	1.00e-11
R16	SiH3 + SiH2 → Si2H5	3.77e-13
R17	SiH2 + H → SiH + H2	7.96e-13
R18	SiH2 + H → SiH3	1.11e-12
R19	SiH2 + H2 → SiH4	2.00e-13
R20	SiH2 + SiH → Si2H3	7.22e-13
R21	SiH + H2 → SiH3	1.98e-12
R22	Si2H6 + H → SiH4 + SiH3	7.16e-12
R23	Si2H6 + H → Si2H5 + H2	1.43e-11
R24	Si2H6 + SiH3 → Si2H5 + SiH4	3.27e-12
R25	Si2H6 + SiH2 → Si3H8	1.20e-10
R26	Si2H6 + SiH → Si3H7	1.00e-11
R27	Si2H6 + Si2H4 → Si4H10	1.00e-10
R28	Si2H6* + M → Si2H6 + M	1.00e-10
R29	Si2H6* → Si2H4 + H2	5.00e6 c-1
R30	Si2H6** → SiH4 + SiH2	2.30e7 c-1
R31	Si2H6** → Si2H4 + H2	2.30e7 c-1
R32	Si2H6** + M → Si2H6 + M	2.00e-10
R33	Si2H5 + H → Si2H4 + H2	1.00e-10
R34	Si2H5 + SiH4 → Si2H6 + SiH3	5.00e-13
R35	Si2H5 + SiH3 → Si2H4 + SiH4	1.00e-10
R36	Si2H5 + SiH3 → Si3H8	1.00e-11
R37	Si2H5 + Si2H5 → Si4H10	1.00e-11
R38	Si2H4 + H2 → SiH4 + SiH2	1.00e-10
R39	Si2H4 + H2 → Si2H6	5.33e-13

R40	Si2H4 + SiH4 → Si3H8	1.00e-10
R41	Si2H3 + H2 → Si2H5	1.70e-12
R42	Si3H8 + H → Si2H5 + SiH4	2.17e-11
R43	Si3H8 + SiH3 → Si3H7 + SiH4	3.30e-12
R44	Si3H8 + SiH2 → Si4H10	1.00e-11
R45	Si3H8 + SiH → Si4H9	1.00e-11
R46	Si3H8 + Si2H4 → Si5H12	1.00e-10
R47	Si3H7 + H → Si3H8	1.00e-11
R48	Si3H7 + SiH3 → Si3H6 + SiH4	3.30e-12
R49	Si3H7 + SiH3 → Si4H10	1.00e-11
R50	Si3H6 + SiH3 → Si4H9	2.00e-11
R51	Si4H10 + SiH3 → Si4H9 + SiH4	3.30e-12
R52	Si4H10 + SiH2 → Si5H12	1.00e-11
R53	Si4H10 + SiH → Si5H11	1.00e-11
R54	Si4H9 + H → Si4H10	1.00e-11
R55	Si4H9 + SiH3 → Si4H8 + SiH4	3.30e-12
R56	Si4H9 + SiH3 → Si5H12	1.00e-11
R57	Si4H9 + SiH2 → Si5H11	1.00e-11
R58	Si4H8 + SiH3 → Si5H11	2.00e-11
R59	Si5H11 + H → Si5H12	1.00e-11
R60	Si5H11 + SiH3 → Si5H10 + SiH4	3.30e-12
R61	Si5H12 + SiH3 → Si5H11 + SiH4	3.30e-12

The influence of discharge power is illustrated in Fig. 3.8. One can see that increasing the discharge power (and consequently the electron concentrations) increases the deposition rate. Moreover, the relative contribution of higher silanes (Si_3H_7) into the deposition grows with increasing power.

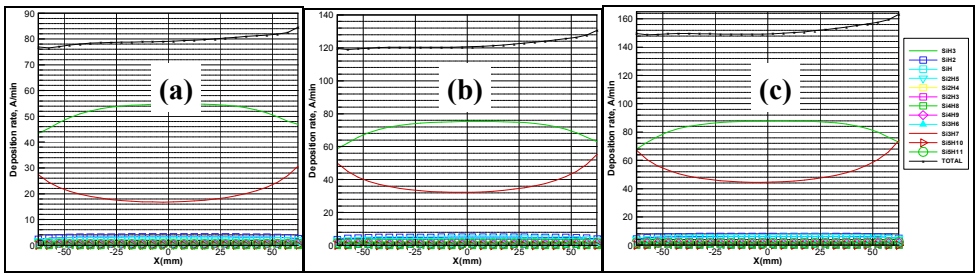


Fig. 3.8. Influence of discharge power *W* on the deposition rate. *p*=0.5 Torr. *W*=0.05 W/cm^2 (a): *W*=0.025 W/cm^2; (b): *W*=0.05 W/cm^2; (c): *W*=0.075 W/cm^2

3.5.3. Influence of reactor geometry

Here, the influence of reactor geometry and the pumping path was investigated. In the first case (Fig. 3.9a), a gas mixture was pumped into the reactor chamber through the

inlets located outside the discharge area, while in the second case (Fig. 3.9b) the mixture enters through a hole in the middle of one of the electrodes (left electrode in Fig. 3.9b).

Fig. 3.9a,b shows the distribution of silyl (SiH_3) and higher silane (Si_3H_8) concentrations. Silyl is the main component that contributes to the film growth. One can see that in the second case (Fig. 3.9a) its distribution along the substrate (that is marked with a bold red line) is more homogeneous, that provides a better homogeneity of the film. However, it grows slightly slower in this case, since the absolute concentration of silyl near the substrate is less. Concentration of Si_3H_8, that triggers dust production, is less in the first case (Fig. 3.9a, right), thus in terms of chemical composition, the film quality is better here.

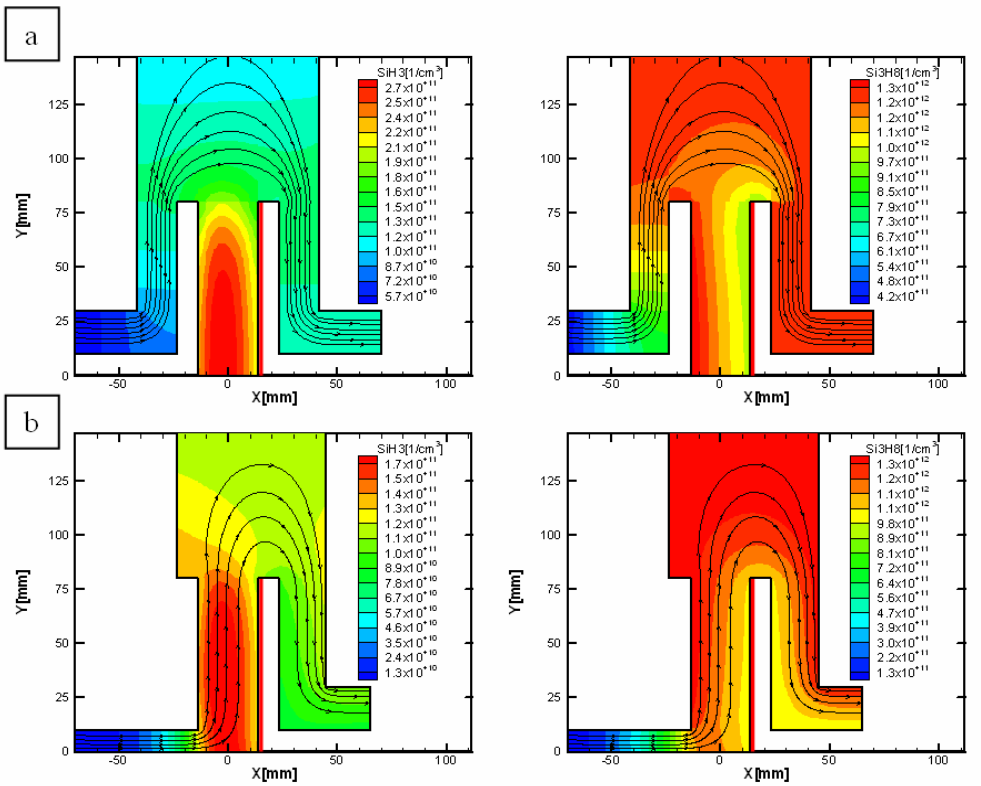


Fig. 3.9. Simulation results: concentration fields of silyl (left), higher silanes (right) and stream lines of the flow (black lines). a, b - different geometrical configuration of reactors. 2D axisymmetric configurations.

Many industrial PECVD reactor configurations are essentially asymmetric and cannot be studied by axisymmetric models. To study the influence of the flow in such chambers, we performed simulations in 2D planar configurations shown in Fig. 3.10. The simulation results showed for instance that adding a small sub-chamber before the pump outlet (Fig. 3.10, B) completely changes the flow pattern: in this case the flow of the source mixture goes from the inlet located at the opposite side *through* the inter-electrode spacing,

whereas without this side-chamber the flow almost does not penetrate the discharge zone (Fig. 3.10, A).

Fig. 3.10. Influence of geometry (location of pumping tube) on the flow behaviour. 2D planar comfigurations.

3.5.4. Deposition dynamics in the final stages of the PECVD cycle

One of the least studied questions in PECVD film deposition is what happens at the end of the production cycle, when the plasma is 'switched off', the inflow of the source gases is stopped, and pressure is increased to atmospheric value. These abrupt changes in the process conditions cause significant changes in the deposition dynamics. Experimental analysis of the processes occurring at this stage is very difficult if not impossible, because the processes of plasma recombination, radicals' diffusion towards the surface and their attachment to the film, are much faster than the time needed for film characterization. Moreover, the measurements of the composition and homogeneity are usually done 'off-line', after the film is taken out of the reactive chamber. In order to study computationally the finishing stages of the PECVD cycle, we carried out a series of simulation experiments varying the basic physical and chemical parameters, as well as the moments in time when these final stages occur (plasma switch-off, gas inflow and heating shut-down) [1]. Some of the results are presented in Fig. 3.11-Fig. 3.13. When the discharge is switched off, we observe not only a rapid decrease in the absolute values of the deposition rate, but also a

strong shift in the film composition, with a noticeable decrease of a relative contribution of silyl and sililyl and increased relative contribution of long-living higher silane radicals (see Fig. 3.11). Another example is stopping the inflow of the source gas mixture into the chamber (with pumping also stopped to keep the pressure constant). It causes not only the obvious decrease in deposition rate due to silane depletion, but also an increased inhomogeneity of film distribution (larger difference between the minimal and maximal deposition rates in Fig. 3.12) and accumulation of higher silanes (see Fig. 3.13), which can form dust particles and stick to the film surface, dramatically changing film properties and causing defects. To avoid this and remove the dusty particles from the reactor volume, a chamber "cleaning" stage is essential with a silane-free gas mixture after the plasma is switched off and before the flow is stopped, without interruption of the production cycle.

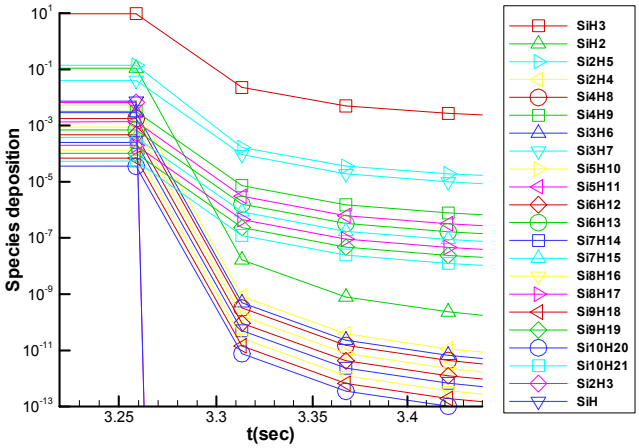


Fig. 3.11. Degradation of film deposition: RF discharge switched off.

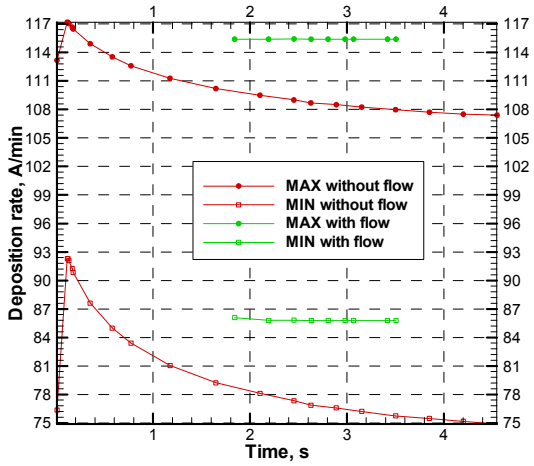


Fig. 3.12. Decrease of the deposition rate after the flow shut-down. Maximum deposition rate in the center and minimal deposition rate at the edges are traced ceparately.

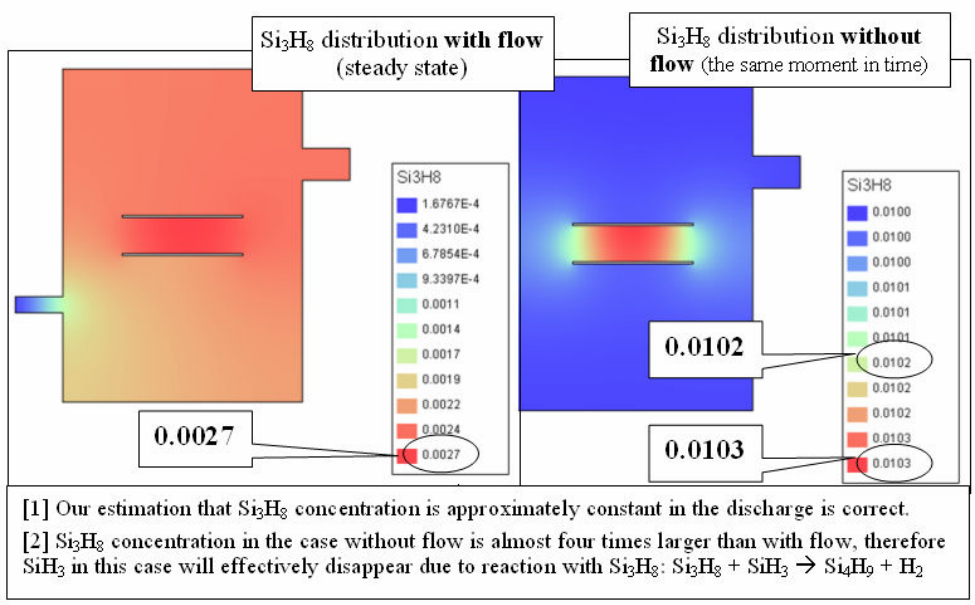


Fig. 3.13. Accumulation of higher silanes in a closed chamber (no inflow and outflow).

3.6. 3D Simulation Results: Flow in Complex Reactor Configurations

The previous section shows that the flow can have a strong influence on the PECVD processes. In industrial reactors this effect can even be more pronounced, due to the essentially asymmetric (3D) geometrical configurations of the reactor chambers. In this section we study the flow in different reactor configurations, without taking into account the plasma and chemical processes.

Four different reactor chamber configurations have been simulated (type A, B, C, D). The models designed accounted for the main features of typical industrial chambers used in plasma doping technology. Small parts such as Langmuir probes and rudimentary side dead-ends were neglected for their effect on the flow dynamics is considered to be insignificant. Fig. 3.14 shows the sketches of the simulated chambers with the generated computational meshes.

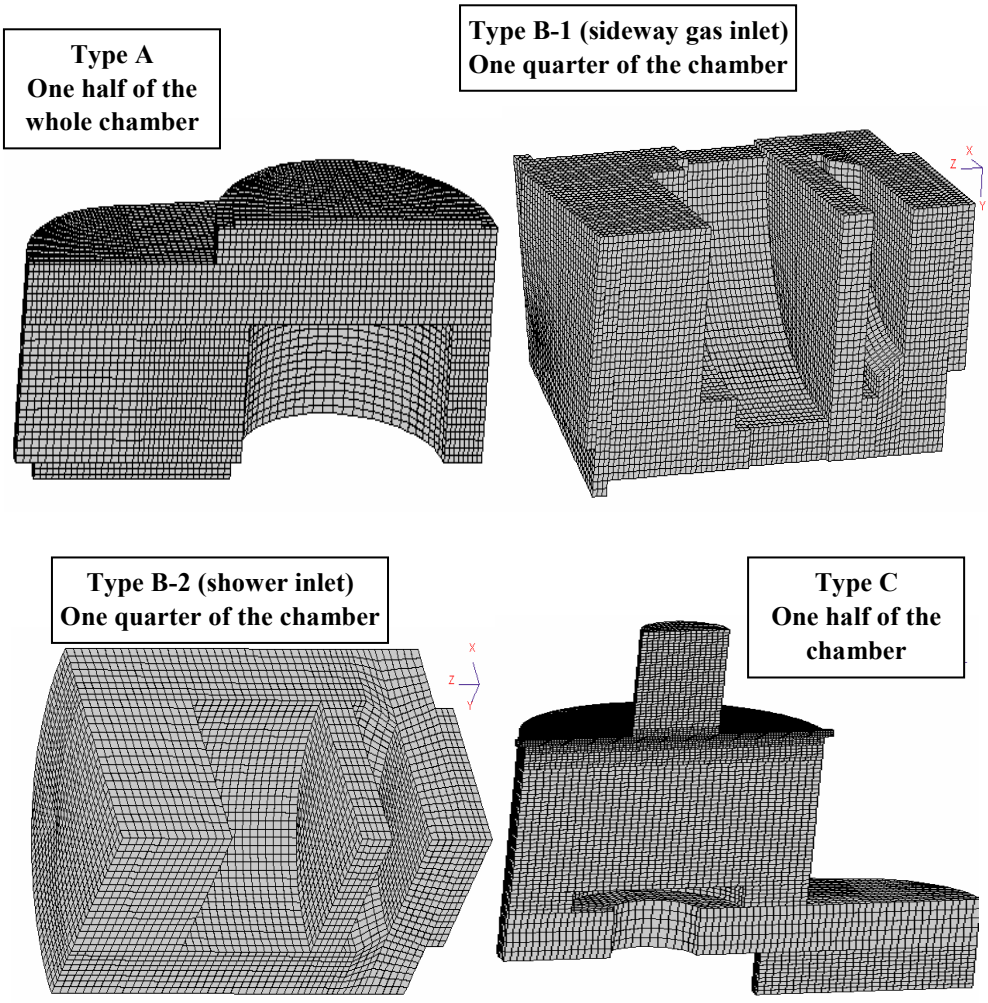


Fig. 3.14. Simulated reactor chambers (4 different models) with the generated computational meshes

Two inflow rates were simulated: 200 sccm and 500 sccm, and three gas mixtures were considered: <1> 0.05% B_2H_6, 99.95% He, <2> 5% B_2H_6, 95% He, and <3> 0.05% B_2H_6, 98.95% He, 1% Ar. The percentage was counted as a mole fraction; therefore for instance 5% of B_2H_6 in case <2> means that the mass fraction of B_2H_6 is 25.5%.

The pressure was fixed to 0.9 Pa, the mixture in the inflow and the initial gas mixture inside the chamber were assumed to be at room temperature (298 K), and the substrate temperature was fixed to 100 deg C (373 K). Sticking coefficient s_i in the boundary condition for species concentration, see eq. (2.11), was assumed to be 1.0 for B_2H_6 on the substrate surface (all molecules that reach this wall stick to it), and 0.0 for other species. This is a very rough model that helps understanding the major trends in species distribution

without going into detail of simulating plasma-chemistry that causes diborane decomposition and sticking of individual radical species.

In the sequel we will only present the simulation results for the A-type chamber at 200 sccm inflow rate and mixture composition case <2>. The results for the other 3 types showed similar trends (data not shown). Fig. 3.15 clearly shows the zone with lower density of the gas above the "hot" substrate. Gas temperature distribution is illustrated in Fig. 3.16.

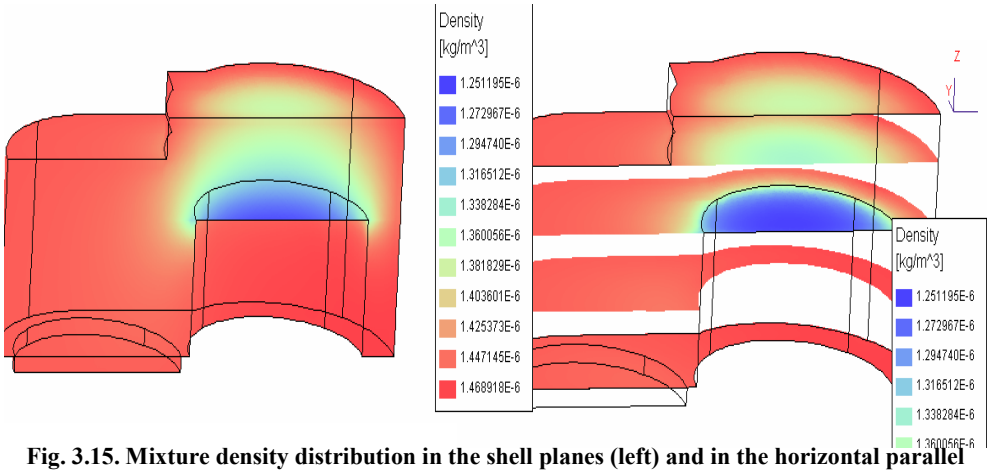


Fig. 3.15. Mixture density distribution in the shell planes (left) and in the horizontal parallel cross-section slices (right).

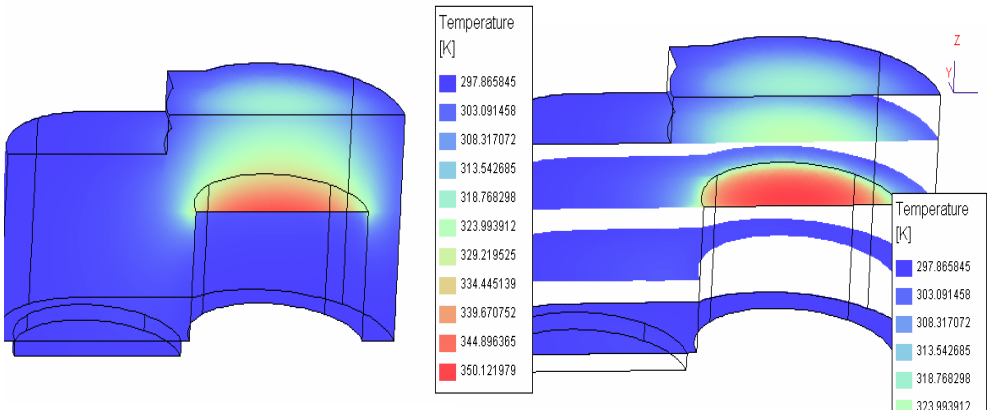


Fig. 3.16. Gas temperature distribution in the shell planes (left) and in the horizontal parallel cross-section slices (right).

The specificity of the chamber design causes high velocities in the area with a relatively narrow flow way between the electrode cylinder and the outer chamber walls. This effect is shown in Fig. 3.17 and Fig. 3.18 with the absolute values and vectors of the flow velocity. Due to this effect, a region with high vorticity appears right behind the edge of the electrode downstream on the way of the main flow path (see Fig. 3.19).

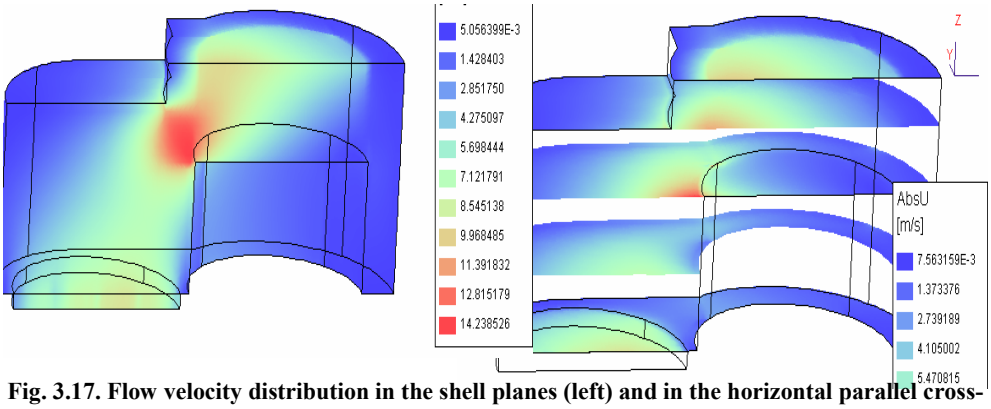


Fig. 3.17. Flow velocity distribution in the shell planes (left) and in the horizontal parallel cross-section slices (right).

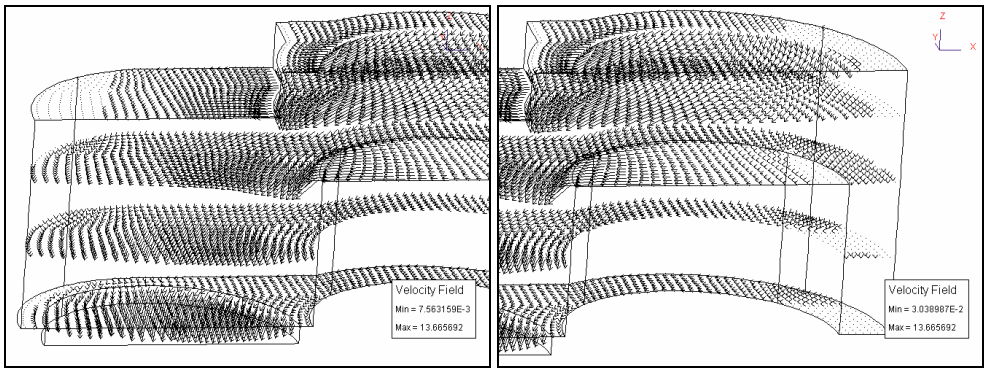


Fig. 3.18. Flow velocity vectors in the horizontal parallel cross-section slices, left and right parts of the chamber.

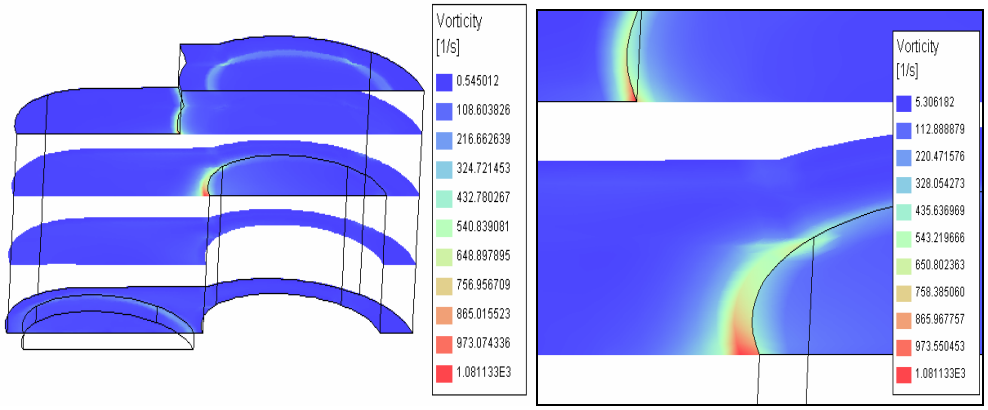


Fig. 3.19. Vorticity distribution in the horizontal parallel cross-section slices, whole chamber (left) and zoom-in in the area with high vorticity.

A detailed study of the dynamics of the spatial distribution of chemical components revealed that convection does not influence the homogeneity of B_2H_6 concentration along the substrate at a 200 sccm flow rate, but causes a slight asymmetry at 500 sccm. As can be seen from Fig. 3.20, diffusion plays a major role in the dynamics of the diborane concentration, which swiftly smoothes throughout the reactor chamber. It is worth mentioning that the concentration of B_2H_6 in the cross-section slice far above the substrate (and close to the inlet) becomes 5 times lower at the time step N=100000 (see the last picture in Fig. 3.20) compared to the initial value (see the first picture in Fig. 3.20). This shows how efficient the diffusion transport is for the conditions simulated.

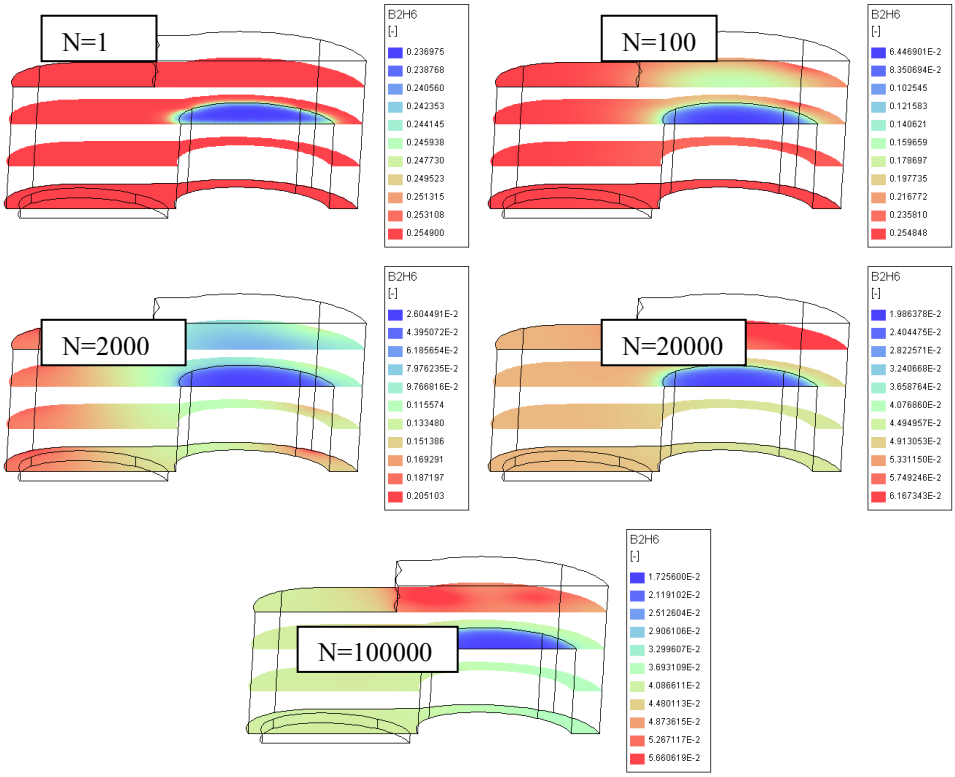


Fig. 3.20. B_2H_6 mass fraction distribution in the horizontal parallel cross-section slices at different time steps N.

These experiments indicate that the homogeneity of the species spatial distribution is mainly defined by diffusive transport and not by convection; therefore the inflow rate can be reduced to the minimal required values just to supply enough of source mixture to the reactive chamber.

3.6.1. Comparison of 3D and 2D Simulation of Gas Flow

To explore a wide range of parameters influencing the flow process and to tune the model, a very large number of simulation runs shall be performed. Full 3D simulations are very time consuming, so we decided to perform 2D simulations for faster analysis. The results of 3D simulations were very encouraging in this respect: there was basically no asymmetry in the radial distribution of the main variables in the geometry of the type B chamber, and the contribution of the third dimension in the type A and Type C configurations was not significant.

To test the accuracy of 2D simulations we chose the Type A chamber geometry. The Type B configuration is almost axisymmetric, and the Type C features a relatively homogeneous flow radial distribution near the wafer because the inflow is organized far from the substrate. Type A is the most asymmetric in the part near the wafer, because the gas enters the chamber relatively close to the electrode.

The configuration of the 2D chamber design was taken equivalent to the geometry of the main plane of symmetry slice used in 3D simulations to cut one half of the computational domain. Fig. 3.21 shows the sketches of the simulated 3D and 2D chambers with the generated computational meshes.

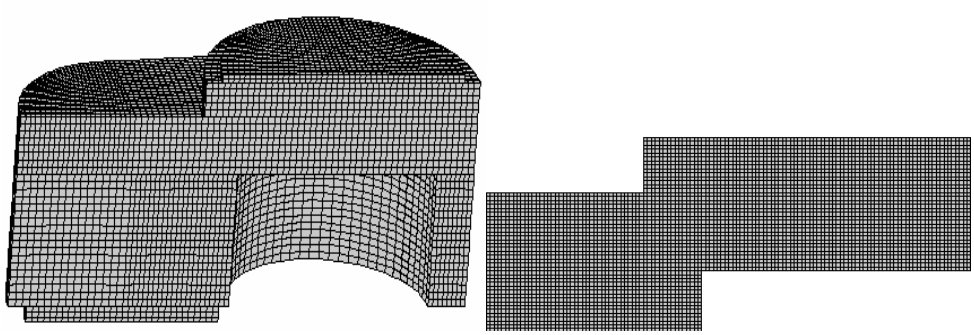


Fig. 3.21. Simulated reactor chambers and the generated computational meshes for the 3D (left) and 2D (right) models of the Type A configuration

Initial and boundary conditions for 2D simulations were the same as for the 3D simulations. Below we show comparison of some of the 2D reactor chamber simulation results with those of 3D configuration. Fig. 3.22 and Fig. 3.23 show the distributions of gas mixture density and temperature, respectively.

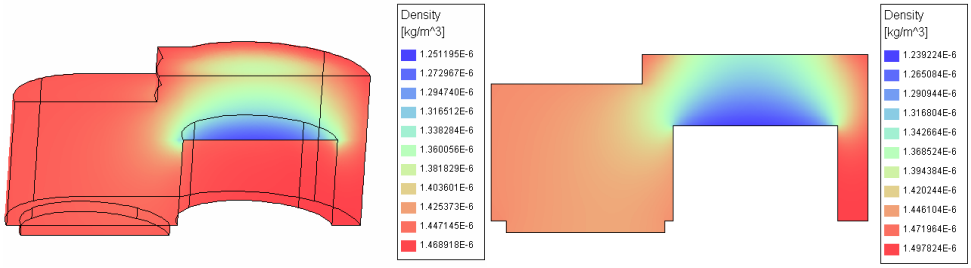


Fig. 3.22. Gas mixture density distribution for the 3D (left) and 2D (right) simulations

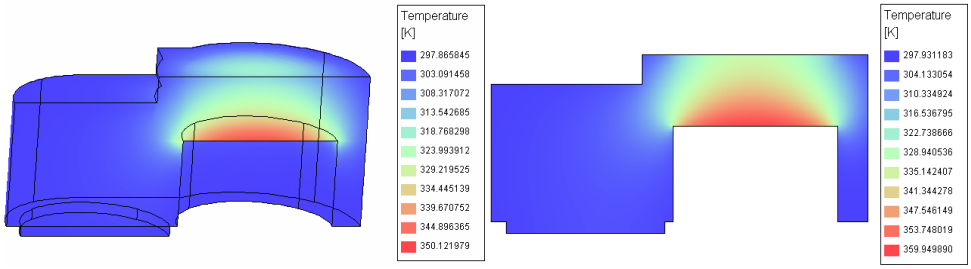


Fig. 3.23. Gas temperature distribution for the 3D (left) and 2D (right) simulations

These spatial distributions in the 2D configuration show nearly no difference compared to those in the median plane of the 3D chamber. A slight difference in absolute value of the maximum temperature in Fig. 3.23 is explained by a different mesh size in 2D and 3D simulations. The visualizer shows the values of the variables in the cell centers, thus the distance from the wall to the center of the first computational cell influences the way the fields are visualized, which is especially noticeable near the heated walls with large parameter gradients.

Fig. 3.24 shows the comparison of the flow velocity distributions. As it was noted earlier, the specificity of the chamber design causes high flow rates in the area with a relatively narrow flow way between the electrode cylinder and the outer chamber walls. This effect is even more pronounced in the 2D simulation since the area, which the current has to squeeze through, is smaller than in 3D simulation. Except for this difference, the overall picture of the flow stays very similar to that predicted by the 3D simulation.

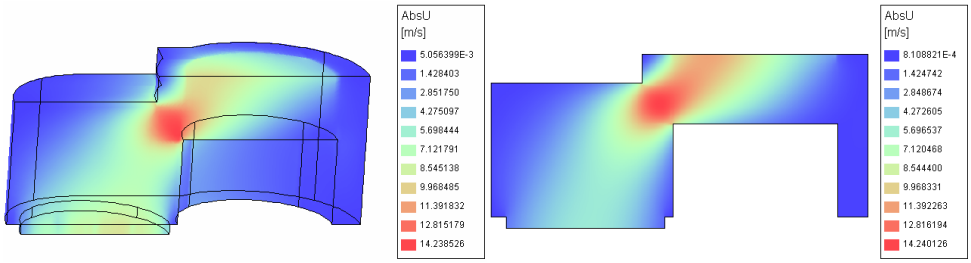


Fig. 3.24. Flow velocity distribution for the 3D (left) and 2D (right) simulations

The same goes for the spatial distribution of B_2H_6 predicted by 2D and 3D simulations. As it can be seen from Fig. 3.25, two-dimensional simulation only slightly underestimates the concentration of diborane in the right lower “sleeve” of the camber, which in 3D is connected to the main chamber through a ring around the electrode, while in 2D it is blockaded. Taking into consideration this fact, we stress once again that this

negligible difference confirms our conclusion that diffusion plays the major role in the dynamics of diborane concentration under the conditions simulated.

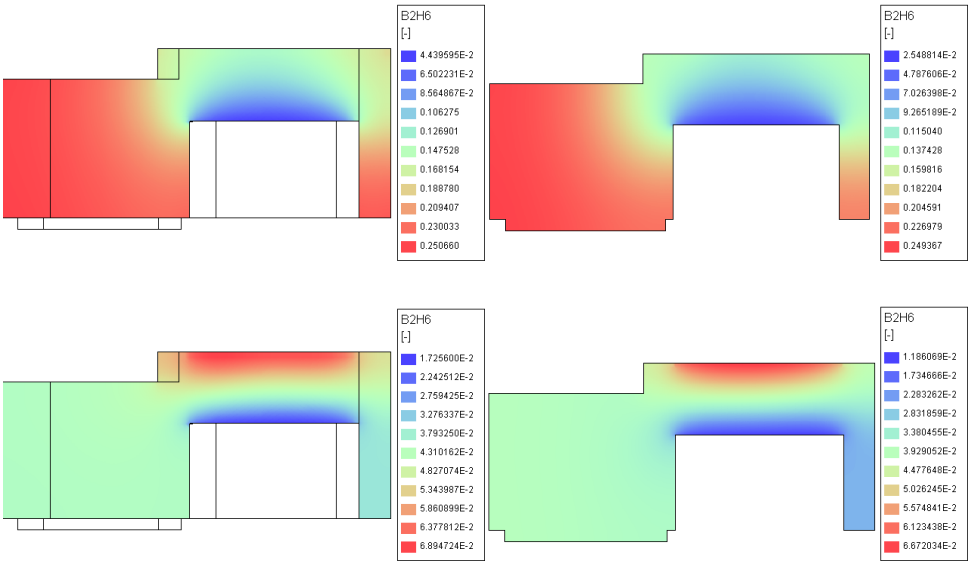


Fig. 3.25. B_2H_6 mass fraction distribution for the 3D (left) and 2D (right) simulations at two different time steps.

Fig. 3.25 also shows that the distribution of B_2H_6 along the wafer near the wall can be considered symmetrical in both 2D and 3D cases, thus bringing us to another conclusion, that the gas dynamic transport cannot significantly influence the characteristics of homogeneity of the doping process. We expect that it will be mainly affected by the plasma discharge distribution and electro-magnetic fields not simulated in the framework of this project.

3.7. Conclusions

Comparison of the results of 2D and 3D simulations shows that all the main characteristics of the flow and, more importantly, the spatial distribution of the chemical species are well captured by the 2D simulations, thus for series simulations and for studying trends it is more efficient to use 2D modeling. Both 3D and 2D simulations showed that the radial distribution of the main parameters affecting the doping process (temperature and species concentrations) along the wafer is almost ideally symmetrical with the axis of symmetry in the center of the wafer. It indicates that the homogeneity of these parameters near the wafer is mainly defined by diffusive transport and not by convection.

Since the gas dynamic transport cannot significantly influence the characteristics of homogeneity of the doping process, the latter will be affected mostly by the plasma discharge distribution and electro-magnetic fields, which were not studied in this chapter.

The Virtual Reactor allowed us to study the film degradation processes and to find the featuring trends in system response to the variation of physical and chemical parameters and the operating actions. We were able to indicate the mechanisms responsible for the layer composition at the finishing stage of the film production cycle.

We have observed a mismatch between the experimental and simulated distribution of the film thickness along the substrate. Additional computational experiments revealed that the correct (experimental) shape can be recovered by adjusting the electron density distribution rather than gas flow parameters like temperature. We therefore conclude that a 2D *discharge* simulation is needed. This is discussed in the next chapter.

3.8. References

1. V.V. Krzhizhanovskaya, P.M.A. Sloot and Yu.E. Gorbachev. Modeling of Plasma Chemical Deposition and Degradation of Silicon Thin Films. The 16th European Conference of Fracture, Alexandroupolis, Greece, July 3-7, 2006.
2. Gorbachev Yu.E., Zatevakhin M.A., Ignatiev A.A., Krzhizhanovskaya V.V. Simulation of heat and mass transfer processes in PECVD reactors for silicon films growth under the conditions of intense formation of higher silanes in the gas phase. The 5th Minsk International Heat and Mass Transfer Forum, MIF-2004. May 24-28, 2004. Minsk, Belarus. Book of abstracts, V. 1, pp. 301-302 (in Russian). Proceedings published on CD, index S04/4-08. Proceedings on the Web: http://relay.itmo.by/forum/mif5/S04/4-08.PDF (in Russian)
3. Gorbachev Yu.E., Zatevakhin M.A., Krzhizhanovskaya V.V., Ignatiev A.A. A Software-hardware System for Simulation of Technological Devices on Distributed Computer Clusters. Proceedings of the X All-Russian conference Telematika-2003. April 14-17, 2003, St. Petersburg, Russia. Publ.: SPbSU ITMO "Informika", Moscow, v. 1, pp. 165-167 (in Russian)
4. V.V. Krzhizhanovskaya, M.A. Zatevakhin, A.A. Ignatiev, Y.E. Gorbachev, P.M.A. Sloot. Distributed Simulation of Silicon-Based Film Growth. Proceedings of the Fourth International Conference on Parallel Processing and Applied Mathematics (PPAM 2001), Lecture Notes in Computer Science, V. 2328, pp. 879-888. Springer-Verlag 2002
5. Y.E. Gorbachev, M.A. Zatevakhin, V.V. Krzhizhanovskaya, A.A. Ignatiev, V.Kh. Protopopov, N.V. Sokolova, A.B. Witenberg. Distributed Simulation of Amorphous Hydrogenated Silicon Films: Numerical Experiments on a Linux Based Computing Environment. Proceedings of tehe International Conference on Computational Science (ICCS 2001), San Francisco, USA, May 2001, in series Lecture Notes in Computer Science, Vol. 2073, pp. 483-491. Springer-Verlag 2001
6. M.A. Zatevakhin, A.A. Ignatiev, V.V. Krzhizhanovskaya, V.Kh. Protopopov, A.B. Witenberg, Y.E. Gorbachev, V.A. Schweigert. Simulation of Heat and Mass Transfer Processes in PECVD Reactors for a-Si:H Films Growth. Proceedings of the XI International Conference on Computational Mechanics and Modern Applied Software Systems. 2-6 July 2001, Moscow, Russia, Pbl.: MAI, pp. 191-193

7. V. V. Krzhizhanovskaya, P.M.A. Sloot and Yu. E. Gorbachev. Grid-based Simulation of Industrial Thin-Film Production. Simulation: Transactions of the Society for Modeling and Simulation International, Special Issue on Applications of Parallel and Distributed Simulation in Industry, January 2005, V. 81, No. 1, pp. 77 - 85
8. V.V. Krzhizhanovskaya, M.A. Zatevakhin, A.A. Ignatiev, Yu.E. Gorbachev, W.J. Goedheer and P.M.A. Sloot. *A 3D Virtual Reactor for Simulation of Silicon-Based Film Production.* Proceedings of the 5th International Bi-Annual Symposium on Computational Technologies for Fluid/Thermal/Structural/Chemical Systems with Industrial Applications, ASME/JSME PVP Conference. San Diego, California USA, July 25-29, 2004. ASME PVP-Vol. 491-2, pp. 59-68, PVP2004-3120. ISBN 0-7918-4686-5
9. Gorbachev, Yu.E., Zatevakhin, M.A., Kaganovich, I.D., 1996, "Simulation of The Growth of Hydrogenated Amorphous Silicon Films from an RF Discharge Plasma," Tech. Phys., **41**, N 12, pp. 1247-1258.
10. Gorbachev, Yu.E., Zatevakhin, M.A., Krzhizhanovskaya, V.V., Schweigert, V.A., 2000, "Special Features of the Growth of Hydrogenated Amorphous Silicon in PECVD reactors," Tech. Phys., **45**, N 8, pp.1032-1041.
11. Lapin, Yu.V., Strelets, M.Kh., 1989, Internal flows of gas mixtures, Moscow, Nauka.
12. Hirschfelder, J.O., Curtiss, C.F., Bird, R.B., 1966, Molecular Theory of Gases and Liquids, John Wiley and Sons.
13. N. Herlin, M. Lefebvre, M. Péalat and J. Perrin, J. Phys. Chem. 96, 7063 (1992)
14. O. Leroy, J. Perrin, J. Jolly, M. Péalat and M. Lefebvre, J. Phys. D: Appl. Phys. 30, 499 (1997)
15. I. Ribet, Th. Pot, M. Lefebvre, J.-P. Taran. Hypersonic Boundary Layer Analysis: Comparison of Velocity Measurements and Monte-Carlo Calculations. 8th AIAA/ASME Joint Thermophysics and Heat Transfer Conference, Saint Louis, MO; United States; 24-26 June 2002.
16. Kogan, M.N., 1969, Rarefied Gas Dynamics, Plenum Press, New York.
17. Godunov, S.K., Riabenkii, V.S., 1973, Difference schemes, Moscow, Nauka, (in Russian).
18. Yang J.Y., Hsu C.A., 1992, "High-Resolution, Nonoscillatory Schemes for Unsteady Compressible Flows," AIAA J., 30, N 6, pp. 1570-1575.
19. Casier F., Deconinck H., Ch. Hirsch., 1984, "A class of bidiagonal schemes for solving the Euler Equations," AIAA J., 22, N 11, pp. 1556-1563.
20. Ignatiev A.A., 2001, "Finite-difference scheme for viscous part of the Navier-Stokes equation," Mathematical Modelling, 13, N 8, pp. 107-116 (in Russian).
21. http://www.gtk.org/
22. http://glade.gnome.org/
23. Nienhuis, G.J., Goedheer, W.J., 1999, "Modelling of a large scale reactor for plasma deposition of silicon," Plasma Sources Sci. Technol. **8,** pp. 295-298.

Chapter 4. Modeling and Simulation Part III: 3D Flow and 2/3D Plasma Discharge*

4.1. Introduction

As we concluded in the previous chapter, a 2D or 3D plasma model is essential to capture the spatial inhomogeneity caused by plasma effects. The goal of this Chapter is to combine the 2D plasma model with the 3D flow model introduced in the previous chapter in order to improve the agreement with experiments. The modeling approach for coupling the models stays essentially the same as described in Section 2.1. In the discharge zone, a 2D plasma model is applied and in the entire reactor volume, a 3D reactive flow model is employed (described in Chapter 3).

4.2. 2D Plasma Model

A 2-dimensional (cylindrically symmetric) self-consistent fluid model [3,4] is used for simulation of the RF discharge processes. The continuity equations for electrons, ions and neutrals are solved consistently with the balance equation for the average electron energy and with the Poisson equation for the electric field distribution.

The rates of inelastic electron collisions and the electron transport coefficients are calculated by solving the Boltzmann equation for the electron energy distribution function (EEDF) in the two-term approximation. The EEDF is calculated as a function of the electric field for a given composition of the neutral background density. A look-up table is constructed to obtain the collision rates and electron transport coefficients as functions of the average electron energy, which are used in the fluid model.

While calculating the EEDF, the electric field was assumed to be spatially uniform and stationary. This limitation was compensated in the fluid model by introduction of the spatial and temporal variation of the averaged electron energy.

Further details on this RF discharge model are published in [3,4], including the list of all chemical reactions simulated in the plasma discharge volume (ionization, dissociation, excitation, recombination, attachment).

4.3. Simulation Results

In order to validate the reliability of the model and the usability of the simulation environment we initiated a number of experiments with the ASTER PECVD reactor from the group of Dr. J.K. Rath from Utrecht University [5]. This allowed us to perform detailed studies in a controlled environment. The geometry of the ASTER system is given in Fig.

* Parts of this chapter were published in [1,2]

4.1. It shows the relevant reactor components such as inlet, outlet, electrodes and overall configuration.

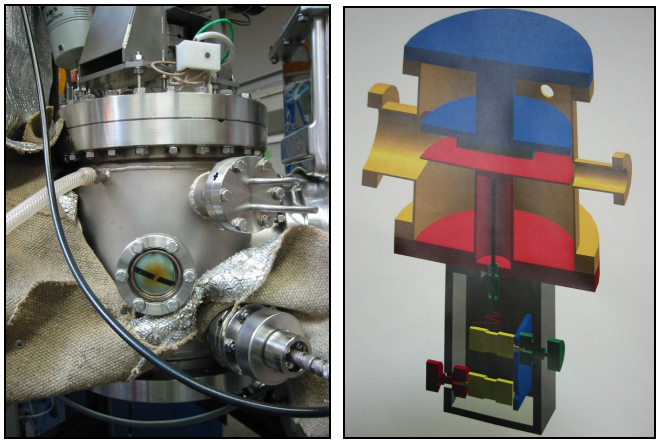


Fig. 4.1. Photo and engineering drawing of the ASTER reactor. The reactor gas chamber has the inner diameter of 200 mm and the height of 280 mm.

We used detailed measurements of the actual geometry of the chamber, substrate and the electrodes as input parameters to our Virtual Reactor. The virtual chamber was constructed from a number of simple blocks, and in each block a regular mesh was generated, as shown in Fig. 4.2.

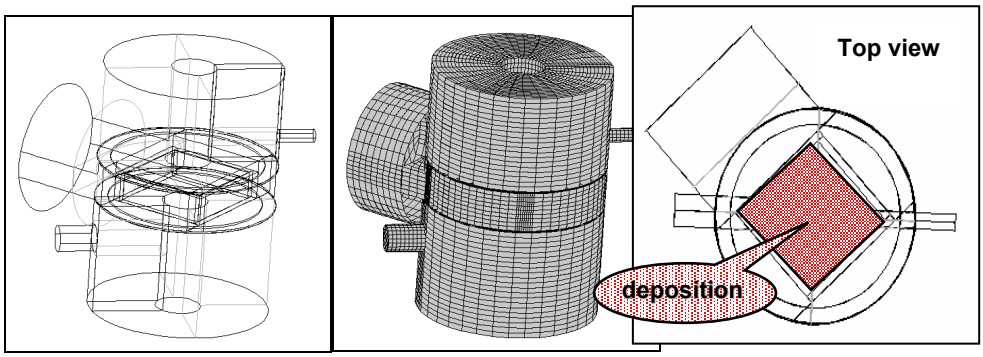


Fig. 4.2. Decomposition of whole computational domain into simple blocks, the generated mesh and the location of the substrate on which the deposition process is taking place.

The initial gas mixture and plasma discharge parameters are chosen corresponding to the real experimental data available [6]. In our comparison the following experimental parameters are used: discharge frequency 50 MHz, applied power 10 W, pressure 0.15 Torr, gas temperature 300 K, substrate temperature 520 K, flow rate 20-60 sccm, inflow mixture composition 50% silane and 50% molecular hydrogen. These conditions were already used in Section 2.7.

Simulation results

We studied in detail the simulation results for the electron distribution as well as the deposition rate on both of the electrodes. In Fig. 4.3 we can clearly see that the area with the maximum electron concentration is located near the edge of the powered (lower) electrode, which corresponds to the edge effects observed in experiments.

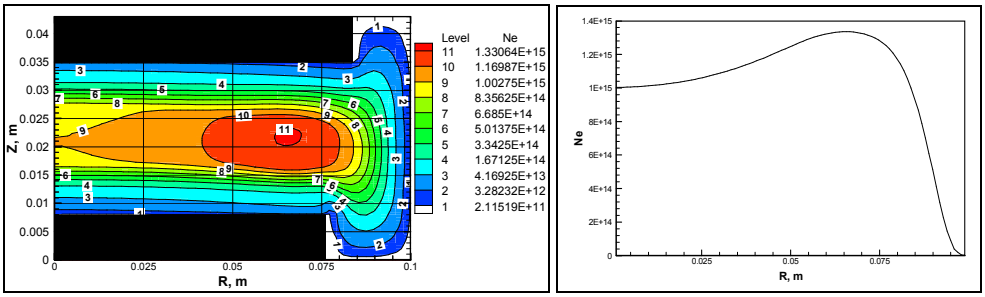


Fig. 4.3. Left: Spatial distribution of the electron concentration. R=0 corresponds to the axis of symmetry. Ne is the number of electrons per m3. Black plates denote electrodes. Right: electron concentration in the horizontal slice at Z=0.021 m (in the middle between the electrodes).

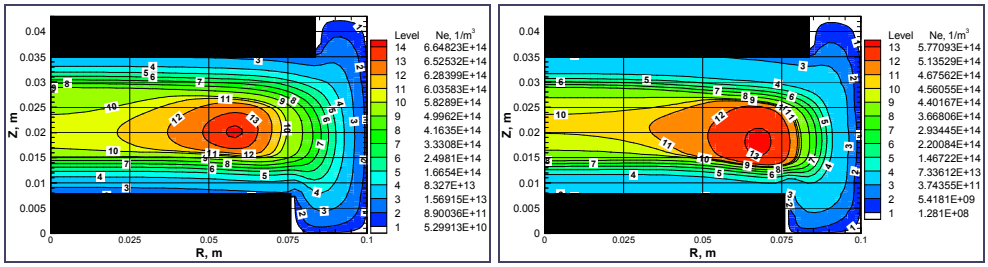


Fig. 4.4. Spatial distribution of electron concentration averaged over RF cycle for different pressures.

We analyzed the spatial distribution of the film thickness. The experimentally observed films have a parabolic variation in thickness, with the minimum in the middle of the substrate, this corresponds to the experimental results (Fig. 4.5).

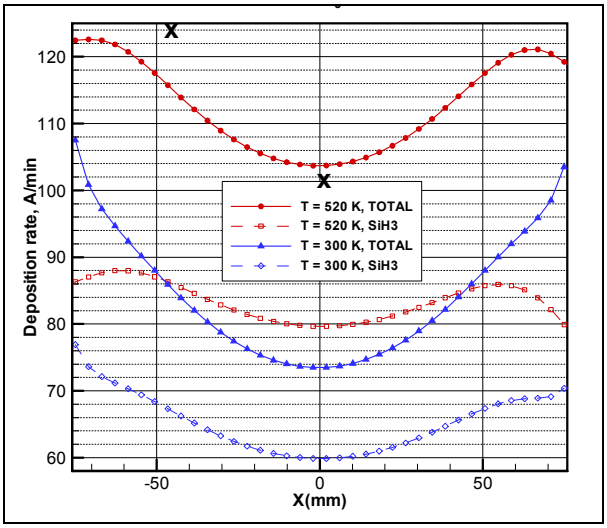


Fig. 4.5. Deposition rate onto the heated (upper) electrode and the 'cold' (lower) electrode. The two experimentally measured values are shown as big crosses 'X'.

4.4. Conclusions

The simulation results are qualitatively in good agreement with respect to the film distribution shape and the observed thickness variation, this variation is more pronounced in the experiment than in the simulation results. In addition we studied the influence of pressure, temperature and plasma discharge parameters (power and frequency) on the PECVD processes in a wide range of parameters. We observed a qualitatively good agreement of the simulated results with the experimental data, for instance the correlation of the measured film properties with the simulated ion and radical fluxes towards the substrate [7,8].

These results clearly indicate that the Virtual Reactor is capable of predicting trends in the most relevant film growth parameters and can be used for studying, predicting and optimizing the PECVD technology.

We also studied the influence of the reactor configuration and flow rate on the deposition processes with commercial software (CFD-ACE [9]). It was shown that small changes in for instance the inlet and outlet position of the flow significantly affect the uniformity and thickness of the growing film. Detailed studies showed the deposition characteristics are sensitive to the complete experimental parameter space (unpublished results, data not shown).

4.5. References

1. V. V. Krzhizhanovskaya, P.M.A. Sloot and Yu. E. Gorbachev. *Grid-based Simulation of Industrial Thin-Film Production*. Simulation: Transactions of the Society for

Modeling and Simulation International, Special Issue on Applications of Parallel and Distributed Simulation in Industry, January 2005, V. 81, No. 1, pp. 77 - 85

2. V.V. Krzhizhanovskaya, M.A. Zatevakhin, A.A. Ignatiev, Yu.E. Gorbachev, W.J. Goedheer and P.M.A. Sloot. *A 3D Virtual Reactor for Simulation of Silicon-Based Film Production.* Proceedings of the 5th International Bi-Annual Symposium on Computational Technologies for Fluid/Thermal/Structural/Chemical Systems with Industrial Applications, ASME/JSME PVP Conference. San Diego, California USA, July 25-29, 2004. ASME PVP-Vol. 491-2, pp. 59-68, PVP2004-3120. ISBN 0-7918-4686-5
3. Nienhuis, G.J., Goedheer, W.J., 1999, "Modelling of a large scale reactor for plasma deposition of silicon," Plasma Sources Sci. Technol. **8,** pp. 295-298.
4. Nienhuis, G.J., 1998, *Plasma models for silicon deposition*. PhD Thesis, Utrecht University.
5. http://www1.phys.uu.nl/SID/programs/physics%20of%20devices/physdev.html
6. Van Sark, W.G.J.H.M., 2001, "Methods of deposition of hydrogenated amorphous silicon for Device applications," in Handbook of Thin Films Materials, Vol. 1: Deposition and Processing of Thin Films, Academic Press, San Diego, CA, USA, pp. 1-102.
7. J.K. Rath, A. Verkerk, M. Brinza, R.E.I. Schropp, W.J. Goedheer, V.V. Krzhizhanovskaya, Y.E. Gorbachev, K.E. Orlov, E.M. Khilkevitch, A.S. Smirnov. Gas phase considerations for the deposition of thin film silicon cells by VHF-PECVD at low substrate temperatures. Accepted at the 33rd IEEE PVSEC.
8. A. Verkerk, J.K. Rath, M. Brinza, R.E.I. Schropp, W.J. Goedheer, V.V. Krzhizhanovskaya, Y.E. Gorbachev, K.E. Orlov, E.M.Khilkevitch, A.S. Smirnov. Compensation of decreased ion energy by increased hydrogen dilution in plasma deposition of thin film silicon solar cells at low substrate temperatures. Accepted at the Symposium K of the EMRS spring meeting.
9. http://www.esi-group.com/products/Fluid-Dynamics/cfd-ace

Chapter 5. Parallelization and Adaptive Load Balancing on the Grid*

5.1. Introduction

Porting complex distributed applications to distributed resources like the Grid poses a grand challenge to the computer and computational sciences, mostly due to the dynamical and decentralized nature of the Grid. Involving parallel computational solvers further complicates the problem because of the severe heterogeneity of Grid resources characterized by a wide range of processors and network communications performance. Lately, the scientific community has been investing lots of efforts into development of Grid-aware problem solving environments (PSE) for complex applications [2,3]. In the next Chapter (Chapter 6), we will discuss the integration of the Virtual Reactor into such a PSE. In this Chapter we address the challenge to efficiently execute the PECVD simulations on the Grid.

Grid computing technologies opened up new opportunities to access virtually unlimited computational resources, and inspired many researchers to work on adaptation of parallel methods and to develop new mechanisms for distributed applications on the Grid. The PECVD Virtual Reactor discussed in this thesis serves as a test-case driving and validating the development of the Russian-Dutch computational Grid (RDG) for distributed high performance simulation [4].

The Virtual Reactor is an excellent example of the multiphysics applications spanning a wide range of spatial and temporal scales [5,6]. Simulation of three-dimensional flow with chemical reactions and plasma discharge in complex geometries is one of the most challenging and demanding problems in computational science and engineering, requiring both high-performance and high-throughput computing.

The Russian-Dutch Grid provides a real-life hardware background for this research as it contains sites with both homogeneous and heterogeneous computing and networking resources. To execute a Grid-enabled complex simulation, a proper functional decomposition of the constituting components is required. To assure that the components, especially computational modules, are distributed efficiently it is necessary to carry out performance evaluation of the individual modules on Grid resources and draw generic conclusions on their behavior: how scalable they are depending on input data and resources used, what is the possible achievable speedup, how infrastructure properties influence the application performance, etc.

A countless number of parallel applications have been developed for traditional (i.e. static homogeneous) parallel systems. Porting such applications from homogeneous

* Parts of this chapter were published in [1]

computing environments to dynamic heterogeneous computing and networking resources poses a challenge to keep up a high level of application efficiency. To assure efficient utilization of Grid resources, special methods for workload distribution control should be applied. Proper workload optimization methods should take into account two aspects: (1) the application characteristics (e.g. the amount of data transferred between the processes, amount of floating point operations and memory consumption) and (2) the resource characteristics (e.g. processors, network and memory capacities, as well as the level of heterogeneity of the dynamically assigned resources). The method should be computationally inexpensive not to induce a large overhead on application performance. In this chapter we present such a method and evaluate it using one of the core parallel solvers of the Virtual Reactor.

The issue of load balancing in a Grid environment is addressed by a number of research groups. Generally studies on load balancing consider distribution of processes to computational resources on the system/library level with no modifications in the application code [7,8]. Less often, load balancing code is included into the application source-code to improve performance in specific cases [9,10]. Some research projects concern load balancing techniques that use source code transformations to improve the execution of the application [11]. We employ an application-centric approach where the balancing decisions are taken by the application itself. The algorithm that estimates the available resources and suggests the optimal load balancing of a parallel job is generic and can be employed in any parallel application to be executed on heterogeneous resources.

A detailed description of global load optimization approaches for heterogeneous resources and adaptive mesh refinement applications is given in [12,13,14]. However, in [12] and [14] no network link heterogeneity was considered and only static resource estimation (initialization) was performed in [12] and [13]. These two issues are the major challenges of Grid computing: 1) the heterogeneity of the network links can be an order of magnitude higher that of the processing power; and 2) Grid resources are inherently dynamic. Developing our algorithm, we tried to address specifically these two issues. The approaches discussed in [12] and [14] are only valid for batch sequential applications (specifically for the queuing systems and computer cluster schedulers), whereas our effort is directed towards parallel programs utilizing heterogeneous resources.

5.2. Parallelization

In order to fully appreciate the distributed nature of the Grid environment on which the virtual reactor will run, we need a distributed, loosely coupled Single Processor Multiple Data-stream (SPMD) type of parallelization [15]. We investigate various levels of parallelization [16]:

5.2.1. Job-level parallelism

Because of the large number of parameters that influence the film properties and deposition rate (pressure, gas and substrate temperatures, flow velocity, reactor volume and configuration, mixture composition and numerous plasma characteristics), for thorough investigation of the system behavior many tasks with different initial conditions can be run

in parallel on different processors, providing exploration of a large parameter space. This way of job-level parallelization (also known as Capacity computing) is the most natural and efficient approach for many computational fluid dynamic problems that require significant efforts for parallelization of individual program algorithms and codes. To utilize the advantages of this approach, a task Manager system has was developed, looking after cluster processor idleness and loading them with new jobs. When one of the processors finished its job, the Manager initiates a process with new initial data set, running over the values of parameters studied (see Fig. 5.1).

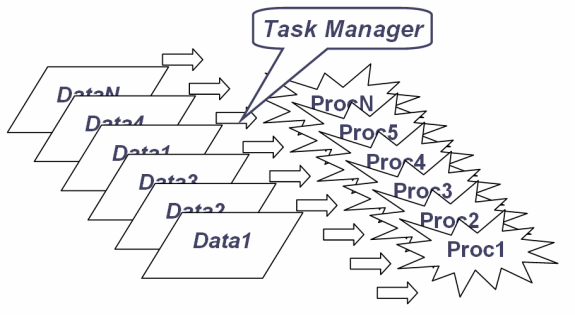


Fig. 5.1 Scheme of job-level parallelization.

5.2.2. Task-level parallelism

Another level of parallelization, for an individual task, is also needed, as one of the aims of this project was the creation of a computational system for large-scale real-time three-dimensional simulation and visualization of complex-shaped industrial reactors.

A natural approach to solve transient computational fluid dynamics problems in parallel is based on decomposition of the computational domain into a number of sub-domains (equal to the number of processors) and on computation of each sub-domain by one of the processors, followed by the data exchange at each time step.

The efficiency of this approach essentially depends on the simplicity of the domain geometry, spatial uniformity of the processes and the communication / computation ratio. A typical spatial domain of a PECVD reactor is usually rather complex, so the trivial geometric decomposition makes sub-domains unbalanced in number of mesh points and total calculations to be done. A specific decomposition algorithm needs to be developed in order to alleviate this imbalance. It is based upon multi-block mesh decomposition with a beam distribution for parallelization. According to this concept the whole simulation domain is divided into a number of primitive blocks. For example, Fig. 5.2 (left) shows a two-dimensional complex domain consisting of nine blocks. Each block is divided into 4-face cells such that the whole domain has two families of beams: in “i” and “j” index directions. Each cell belongs to only one beam of each family and each beam includes a sequence of cells with adjacent faces such that one beam begins and finishes at the physical boundary. An advantage of this beam decomposition approach is that it supports the use of a uniform implicit scheme by applying sweep-type algorithms to the single beam.

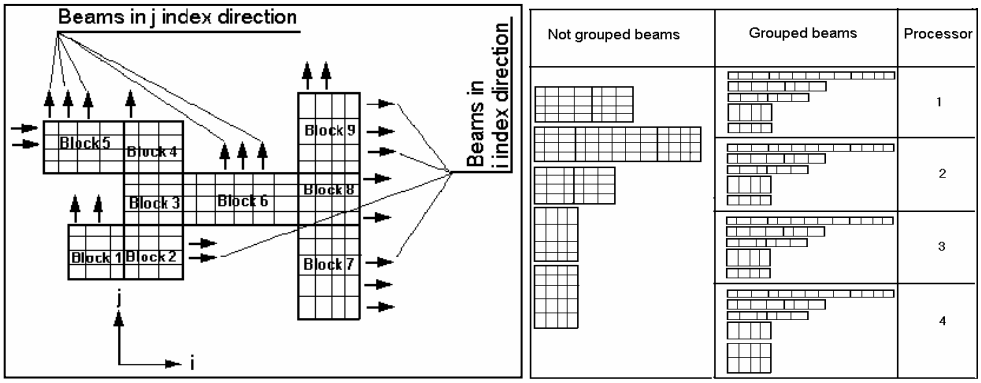


Fig. 5.2 Left: Decomposition algorithm for multi-block meshes and implicit numerical schemes. Right: Beam distribution for parallelization.

The parallel algorithm boils down to the following steps:

• Formation of two double-linked lists of beams in both index directions ("i" list and "j" list) for the given domain.

• Compilation of NP groups of beams inside each of two lists. (NP – number of processors), so that each group has approximately equal number of cells.

Fig. 5.2 (right) illustrates grouping of beams for the "i" list of our nine-block domain from for four processors. One can see that in this case loading of all processors is approximately equal. Similar grouping is made for the "j" list.

• Based on this beam grouping, the work instructions are formulated for each processor, listing the beams to be calculated by every processor.

• According to these instructions, in the beginning of each time step the master processor sends necessary information to slave processors. Each slave processor receives the data and computes the beams listed in its instructions.

• After performing its management duties, the master processor accomplishes its own calculation tasks and waits for results (decrements of conservative variables) from the slave processors. On receiving these results it merges all data arrays into one and repeats the previous step.

This algorithm for 1 time step is briefly illustrated in Fig. 5.3.

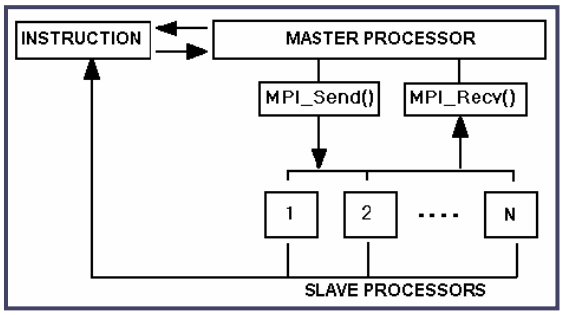


Fig. 5.3 Scheme of task-level parallelization algorithm.

In [16] some performance measurements on the task-level parallelization are shown, indicating the rapid saturation of the parallel efficiency due to the imposed communication overhead. The parallel efficiency was calculated as the ratio of t_{seq} to $n \cdot t_n$, where t_{seq} is the time consumed in a sequential mode and t_n is the computational time for a parallel algorithm to run on n processors. More details on the parallelization strategies and the associated performance results are provided in [16].

5.3. The simulation components of the Virtual Reactor

The Virtual Reactor includes the basic components for reactor geometry design; computational mesh generation; plasma, flow and chemistry simulation; editors of chemical processes and gas properties connected to the corresponding databases; pre- and postprocessors, visualization and archiving modules [5]. The aim of our research is to virtualize separate modules of the application and to run them efficiently as services and access them on the Grid.

The application components perform one (or a few) of the following functions: problem description, simulation, visualization and interaction. This is schematically shown in Fig. 5.4, where we emphasize the *simulation* components.

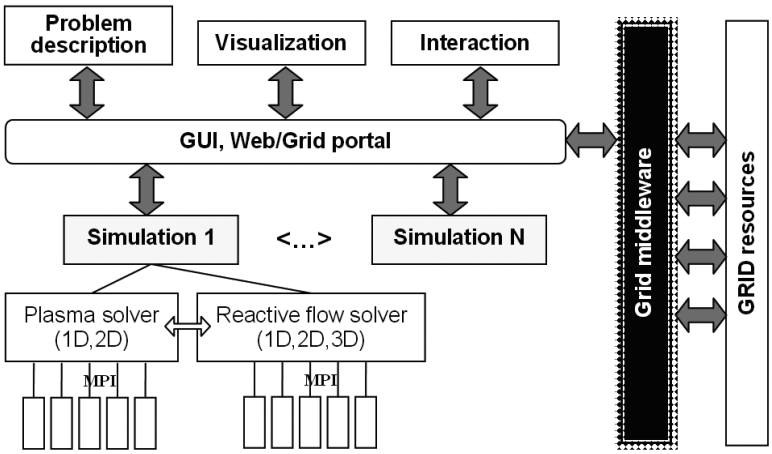


Fig. 5.4 A hierarchy of simulations. Plasma and Reactive flow solvers are the most time-consuming simulation components.

The core components are modules simulating plasma discharge, gas flow, chemical reactions and film deposition processes occurring in a PECVD reactor. The details on numerical methods and parallel algorithms employed in the solvers are described in previous chapters and in [16]. The most important features relevant to the Grid implementation are as follows: for stability reasons, implicit finite volume schemes were applied, thus forcing us to use a sweep-type algorithm for solving equations in every "beam" of computational cells in each spatial direction of the Cartesian mesh. A special parallel algorithm was developed with beams distributed among the processors.

Communications are organized exploiting a Master-Slave model; where at each simulated time step the Master prepares instructions for the Slaves, sends them the data to be processed, receives the results, and processes them before proceeding to the next step. The algorithm is implemented in an SPMD model, using the MPI message passing interface with MPI Barrier points for synchronization. Data exchange between the Master and the Slaves is repeated every time step, and simulation proceeds for thousands to millions of steps. In the testbed we use generic MPICH-P4 built binaries that can be executed on all the testbed machines using the Globus job submission service. To study the influence of various parameters on the simulated processes we run a number of simulations in parallel (shown in Fig. 5.4 as "Simulation 1" … "Simulation N" blocks) with the assistance of Nimrod-G [17].

To provide efficient execution of a parallel application on heterogeneous resources, one needs to clearly understand the application performance dependencies on homogeneous resources first. This gives an insight into the application scalability, induced fractional overhead, dependencies of the amount of the communications and calculations on the number of processors used, etc. The results of such tests can help estimating and predicting the behavior of the application on heterogeneous resources, thus simplifying the adaptation process.

5.4. The Russian-Dutch Grid testbed infrastructure

Generally the infrastructure of a site within a Grid testbed can be of one of the following types depending on the underlying resources:

I. traditional homogeneous computer cluster architecture: homogeneous worker nodes and uniform interconnection links;
II. homogeneous worker nodes with heterogeneous interconnections;
III. heterogeneous worker nodes with uniform interconnections;
IV. heterogeneous nodes with heterogeneous interconnections.

A complete Grid infrastructure is always of the Type IV, characterized by severe heterogeneity with a wide range of processor and network communication parameters. As we show later in this paper, the type of resources allocated to a parallel application significantly influences its performance, and different load balancing techniques shall be applied to different combinations of the resources. Currently the Russian-Dutch Grid testbed consists of six sites with different infrastructures:

Amsterdam-1 (contains 3 nodes, 4 processors) – Type IV;
Amsterdam-2 (32 nodes, 64 processors) – Type I;
St. Petersburg (4 nodes, 6 processors) – Type IV;
Novosibirsk (4 processors) – Type II;
Moscow-1 (13 nodes, 26 processors) – Type I;
Moscow-2 (12 nodes, 24 processors) – Type I.

The Russian-Dutch Grid testbed is built with the CrossGrid middleware [3] based on the LCG-2 distributions and sustains the interoperability with the CrossGrid testbed. More detailed information on the RDG testbed can be found in [4]. The RDG Virtual Organization (VO) is included into the CrossGrid VO, thus allowing the RDG certificate holders to access some of the CrossGrid resources and services. The CrossGrid testbed consists of 16 sites with the infrastructures of all 4 types.

5.5. Performance analysis on homogeneous resources

5.5.1. Benchmark approach

Benchmarking of a complex application is required to evaluate its performance and reveal the dependencies of its behavior on the underlying infrastructure. We use a structural approach to benchmark the Virtual Reactor as an example of a complex application. Within this approach, the overall functionality of the whole system is studied, followed by performance measurements of the individual components while they are not influenced by activities of the other components.

Benchmarking the components of a complex problem solving environment allows evaluating their performance depending on various parameters like types of input data and the resources used. This helps to predict the performance of a given component and use it for efficient resource allocation, thus improving the overall resource management within the whole application.

The earlier tests of the Virtual Reactor performed on the CrossGrid testbed showed that most of the interactive components of the Virtual Reactor do not put restrictions on the computer systems and network bandwidth and can be efficiently executed on distributed Grid resources [5]. Next, we focused on benchmarking of the *simulation* modules. Each simulation consists of two basic components: one for plasma simulation and another for reactive flow simulation (see Fig. 5.4). These two components exchange only a small amount of data every hundred or thousand time steps, therefore the network bandwidth is not a critical factor. Finally, we concentrate on benchmarking the individual parallel solvers, starting from a 2D PECVD solver which contains all the features of the 3D solver but takes less time to estimate the solver behavior on the Grid.

5.5.2. Benchmark setup

The goal of the benchmarking is to determine the scalability of the application, find out the limitations on the efficiency posed by the application architecture, resources and types of the simulations. Uncovering such details will allow us to optimize a resource management strategy for allocating the application components within the whole Virtual Reactor problem solving environment.

The solver operates a reactor geometry that is composed of a number of connected blocks. Different types of simulation can be performed within a single geometry: a chemically inactive flow and a flow with chemical and plasma processes. Physically the problem type is determined by the gas mixture composition, temperatures, pressures, and

the plasma discharge operation mode. From a computational point of view these types of simulations differ by the ratio of computations to communications: for simulating chemical processes the computational load is significantly higher.

We start from a light-weighted problem not simulating the chemical and plasma processes and with a simplified reactor geometry consisting of a single block that allows for easy tracking of parameter influences on the execution time. To measure the dependency of the solver performance upon the input data, multiparameter variation is applied. We measured the solver execution time, speedup and communication time depending on the combinations of input parameters: the computational mesh size, number of simulation time steps and number of processors.

The benchmark tests had to be automated because the parameter variation leads to a large number of job submissions. To solve this problem we have built an execution environment that supports a series of parameter-sweep job submissions in Globus. The environment is generic and can be used for any kind of performance benchmarks with user-defined metrics and parameters to be analyzed. Within this environment, the application to benchmark is described using some templates that are filled with particular application data (e.g. the Globus RSL template for job submission which also contains a list of input and output files). One of the functionalities of this execution environment is the support for parameter-sweep runs, analogous to what Nimrod-G or Condor-G provides. The advantage of our implementation is that we can specify the parameters (and their ranges) that shall be changed, as well as the characteristics to be measured and visualized automatically to analyze the influence of those parameters.

In these tests, a single-block topology was used. The block was subdivided into a (*ncell* x *ncell*) number of computational mesh cells, with ncell running from 40 to 100, thus forming 1600–10000 cells. We performed also some tests with real reactor geometries in order to check whether the reactor topology influences the parallel performance, since potentially it can introduce some load imbalance.

5.5.3. Influence of the number of time steps and reactor topology

Experiments with a different number of time steps showed that the execution time and other measured parameters are linearly proportional to the number of time steps, provided that this number is high enough and the standard output and hard disk operations are kept minimal (that means no excessive logging, nor any storing of the 2D fields or other additional files every time step). All the results presented below are measured for 100 time steps.

Along with the single-block geometry, we studied the performance of the solver with a complex multi-block PECVD reactor topology, which consists of an equivalent number of computational mesh cells. The results showed that all the measured characteristics of the solver behavior (execution time, speedup, computation and communication time) on the same resources do not differ for the single-block and multi-block topologies of equal number of cells within 1% accuracy. This assures us that the parallel algorithm used in the solver provides a good load balancing even in cases of complex topologies. Further we test the influence of the problem size (the number of mesh cells) with the single-block reactor

geometry, since it is easier to vary the mesh size arbitrarily with a single-block geometry than with a multi-block complex topology.

5.5.4. Speedup of the chemistry-disabled and chemistry-enabled simulations

The measurements were carried out on all the Grid sites within the RDG testbed. The parallel solver showed a noticeable speedup on the Moscow and Amsterdam sites of Type I (homogeneous cluster with uniform communication links). Fig. 5.5 and Fig. 5.6 demonstrate the total execution time and speedup of the parallel solver for different types of simulation: A chemistry-disabled "light-weighted" simulation and a chemistry-enabled "heavy" simulation respectively.

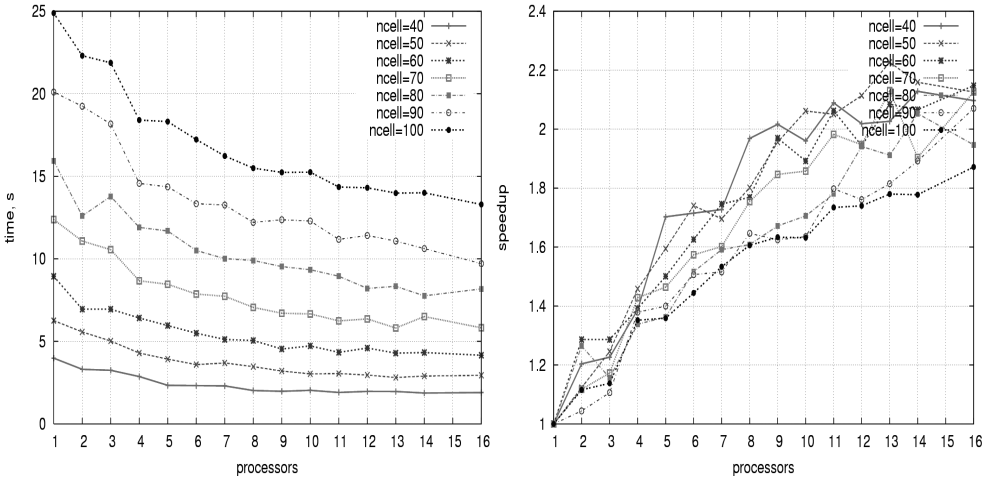


Fig. 5.5. Light-weighted (no chemistry) simulation: total execution time and speedup for different computational mesh sizes.

We observe different trends of the solver performance: for the light-weighted simulation, the speedup decreases with the increase of the mesh size (see the different curves in Fig. 5.5, right), while for the chemistry-enabled simulation, the speedup increases with increasing problem size (Fig. 5.6). The different absolute values of the speedup are mostly dependent on the resources: The results presented in Fig. 5.5 were obtained on the Moscow-1 site with slow inter-processor links, and Fig. 5.6 shows the results of the Amsterdam-2 site with fast communications. The gap between the *ncell*=90 and *ncell*=80, observed in Fig. 5.6 left, is most likely due to non-linear aspect of computations and communications, as well as cash saturation effects. We did not pursue deeper analysis of this result.

The same parallel solver tested on homogeneous Grid sites with a higher ratio of the inter-process communication bandwidth to the processor performance achieved much higher speedups. For instance on lisa.sara.nl with Infiniband interconnections it was 3 times higher for large problem size simulations. The type of MPI library also influences the parallel efficiency of a program: a specialized library optimized for the native communication technology (e.g. MPICH-GM for Myrinet communications on

das2.nikhef.nl) increases the speedup up to two times compared to the generic MPICH-P4 or MPICH-G2.

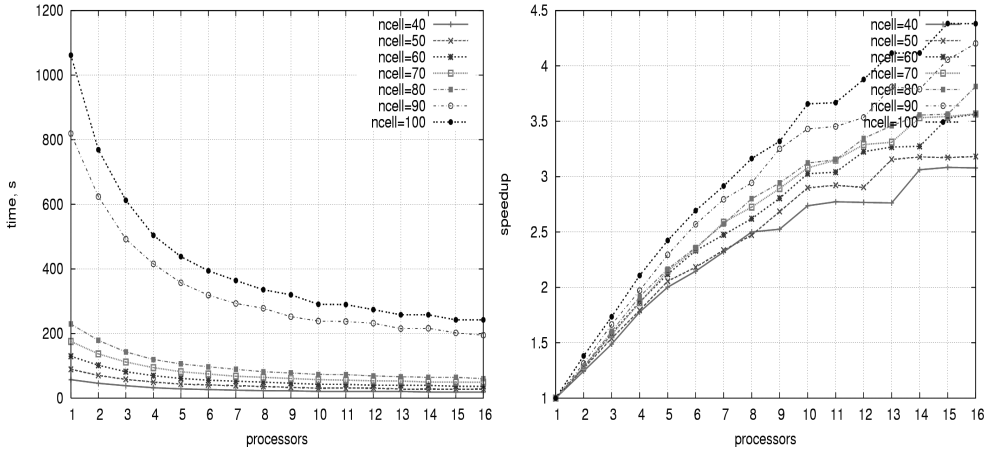


Fig. 5.6. Chemistry-enabled simulation: total execution time and speedup for different computational mesh sizes

5.5.5. Communication time trends

The time spent on inter-process communications within the solver is shown in Fig. 5.7 for different mesh sizes. The communication time was calculated as a sum of MPI Send/MPI Receive time on the master node over the total number of iterations.

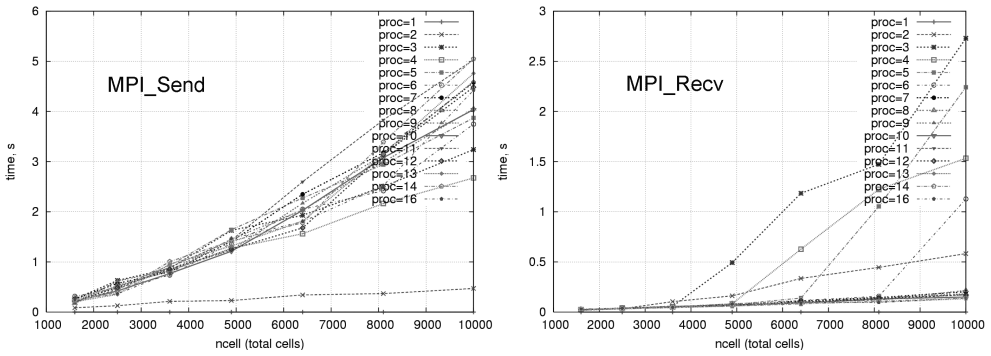


Fig. 5.7. Dependency of the communication time on the computational mesh size for different number of processors (light-weighted simulation)

We observe that communication time grows super-linearly with the increase in mesh size, although the amount of data transferred is linearly proportional to the number of mesh cells. The exact understanding of this behavior is not relevant and falls outside the scope of this thesis. The fact that the time spent for sending the data is not equal to the time of receiving the data in Fig. 5.7 can be explained by the type of measurements: both plots represent the data sent or received by the master node only, as a synchronizing process. As

one can see, the size of the received data (which represent the results of the calculations performed on the slave nodes) is significantly less than the data sent.

Some peculiarities in the communication time can be seen from Fig. 5.7, which are even more explicit in Fig. 5.8: (1) The communication time grows non-monotonically with the number of processors, but drops down a little on every processor with an even number; and (2) The time of MPI Receive calls is an order of magnitude higher for the larger meshes on the first few processors.

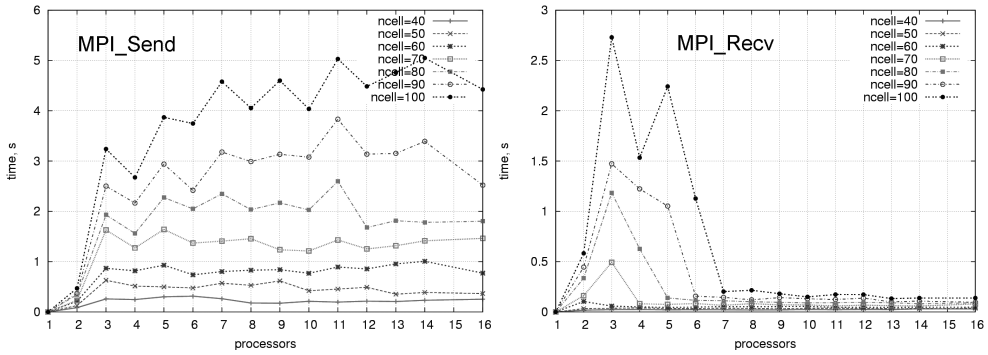


Fig. 5.8. Dependency of the communication time on the number of processors for different computational mesh sizes (light-weighted simulation)

5.5.6. Computation to communication ratio

In Fig. 5.9 the total execution time is presented along with the contributions of calculation and communication. For a smaller computational mesh, the communication time makes a relatively small contribution to the total execution time even for a large number of processors involved. For a larger mesh, the communication makes up to 30% of the execution time. This result confirms that the network bandwidth is not sufficient for this type of problem.

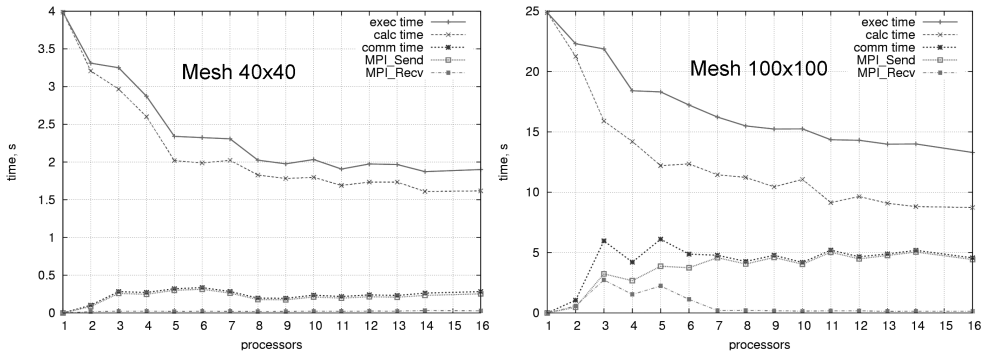


Fig. 5.9. Total execution time and contributions of the calculation and communication depending on the number of processors for different computational mesh sizes (light-weighted simulation)

Fig. 5.10 demonstrates the ratio of computation to communication time for different mesh sizes with different types of the simulation. The higher the ratio, the less communications are required, which obviously offers a better parallel efficiency and application scalability. The ratios in Fig. 5.10 explain the different speedup trends observed for chemistry-enabled and chemistry-disabled (light-weighted) simulations. From the presented graphs we can see that the behavior of this ratio does not depend on the mesh size for the chemistry-enabled simulations, while this behavior for the light-weighted simulations significantly differs for small and large mesh sizes. For a small mesh size, the ratio stays decently high, and for 6 processors and more it reaches the level of the chemistry-enabled simulations. For a larger mesh, the computation/communication ratio for the no-chemistry simulations is very low, thus diminishing the overall parallel efficiency.

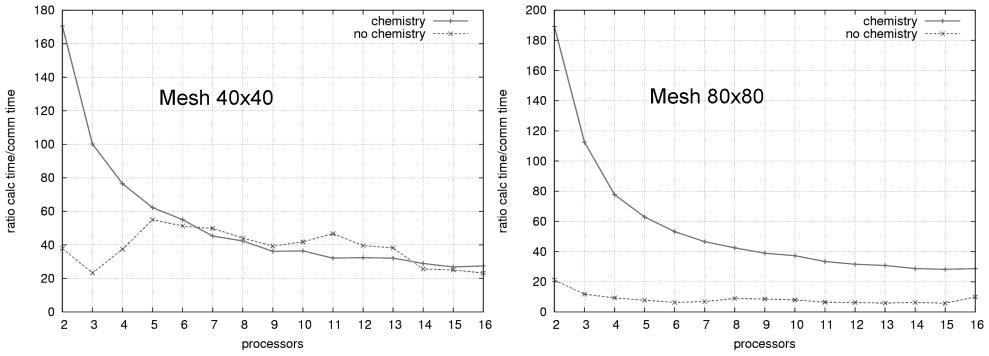


Fig. 5.10. The ratio of the computation to communication time for chemistry-enabled and light-weighted simulations

5.5.7. Discussion of the results for homogeneous resources

The results presented in this section show that the parallel speedup is lower for a larger problem size (with more computational mesh cells) for the simulations of problems without chemical processes. This fact indicates that the ratio of the inter-process communication bandwidth to the processor performance was not high enough for light-weighted problems with relatively small number of operations per computational cell. It means that for optimal usage of the computing power, a large number of processors for one parallel run shall only be used for relatively small computational meshes. Thus the communication technology puts a limit to the scalability of the solver for this problem type. On the other hand, the simulation of the flow *with* chemical processes shows higher speedup with larger meshes (see Fig. 3). Here the amount of computations brought by simulating the chemistry changes the behavior of the solver qualitatively. This leads us to the conclusion that different resource allocation strategies should be applied for different types of simulation and meshes used.

The results in Fig. 5.8 reflect the network and nodes features of the tested Grid site:

1. Since the site consists of dual nodes, the network channels work more efficiently for data transfers between the Master and a Slave processor if a connection was already established with another Slave processor on the same node. This can be explained by

implementation of the MPI library which saves network resources while opening and maintaining connections for concurrent processes on the same node.

2. The "peaks" of the MPI Receive time for the first few processors (see Fig. 5.8, right) are caused by the constraints on the size of the data that could be accommodated at once. The constraining factors can be the network bandwidth distribution, the processor cache size, the memory available on the node, or a combination of these factors.

5.6. Application performance on heterogeneous resources

The RDG Grid sites with heterogeneous processors and/or network links (Types II, III, IV) provided only a limited parallel speedup or even a slow-down of the original solver with a homogeneous parallel algorithm (data not shown). This was inevitable since, in addition to the low-bandwidth links, these sites are characterized by very diverse resources: the processor and network parameters differ by orders of magnitude for different nodes.

The parallel algorithm of the solver was originally developed for homogeneous computer clusters with equal processor power, memory and inter-processor communication bandwidth. In case of submitting equal portions of a parallel job to the nodes with different performance, all the fast processors have to wait at the barrier synchronization point till the slowest ones catch up, thus the effect of slow-down on heterogeneous resources is not surprising. The same problem occurs if the network connection from the Master processor to some of the Slave processors is much slower than to the others. As we have shown in the previous section, for communication-bound simulations (chemistry-disabled simulation with large computational meshes), the communication time on low-bandwidth networks is of the order of the calculation time, therefore the heterogeneity of the inter-processor communication links is a hindrance as considerable as the diversity of the processor power. One of the natural ways to adapt the solver to the *heterogeneous* Grid resources is to distribute the portions of job among the processors according to the processor performance and network connections, *taking into account the application characteristics*. To adapt the parallel solver, we developed a new method presented in the next Section.

5.7. A new methodology for resource-adaptive load balancing on heterogeneous resources

One of the factors that determine the performance of parallel applications on heterogeneous resources is the quality of the workload distribution, e.g. through functional decomposition or domain decomposition. Optimal load distribution is characterized by two things: (1) all processors have a workload proportional to their computational capacity; (2) communications between the processors are minimized. These goals are conflicting since the communication is minimized when all the workload is processed by a single processor and no communication takes place, and distributing the workload inevitably incurs communication overheads. Thus it is needed to find a trade-off and define a metric that characterizes the quality of workload distribution for a parallel problem. One of the existing methods to measure it is to introduce a cost function reflecting the application execution time. Minimization of this function corresponds to minimization of the application runtime.

The function should be simple and independent of the details of the code. The generic form of such a cost function is [18,19,20]:

$$H = H_{calc} + \beta H_{comm}, \tag{5.1}$$

where H_{calc} is minimized when the workload distribution among the processors is proportional to the processors capacity (or equal in case of homogeneous processors); H_{comm} is minimized when the communication time is minimal; and β is a parameter that can be varied in order to tune the balance between the calculation and communication terms. This parameter is dependent on the characteristics of both the application requirements and the resources capabilities.

The main generic parameters that define a parallel application performance are:

- An application parameter $f_c \sim N_{comm}/N_{calc}$ (N_{comm} and N_{calc} are the amounts of application communications and computations respectively);
- A resource parameter $\mu \sim t_{comm}/t_{calc}$ (t_{comm} is a typical time taken to communicate a single word between the processors, t_{calc} - typical time required to perform a generic floating point calculation).

The product of these two parameters $f_c\mu$ is often called the fractional communication overhead [18].

The goal of load balancing is to minimize the cost function (5.1). The parameter β in this expression is an aggregated value based on the application and resource specific parameters f_c and μ. Knowledge of these application and resource properties allows constructing an appropriate form of parameter β and performing suboptimal load distribution [20]. However in most real-life complex simulation problems, it is not possible to theoretically calculate the application specific parameter f_c with a reasonable precision. Even a detailed analysis of the algorithms and codes can fail in many practical cases when the code has multiple logical switches and completely different algorithms and computational schemes are used while solving a problem, depending on the initial conditions and computational parameters. Estimation of the resource-specific parameter μ also poses a challenge on heterogeneous Grid resources, since there is a multitude of processors with the ratio of communication to computation performance spanning a few orders of magnitude. Moreover, the Grid exhibits dynamic network and processor performance, therefore static domain decomposition fails to provide realistic estimations and consequently optimal load distribution. To ensure efficient load balancing of a parallel application on the Grid, it is necessary to estimate the β parameter experimentally. There are two possible approaches to that: (1) directly measure the lumped value of β for the application on the allocated resources and (2) separately benchmark the resources, estimate μ and then find out the application-specific parameter f_c that would provide an optimal workload distribution on a given set of resources. The first approach requires serious intrusion into the application code. This is certainly not desirable, especially when targeting to build a generic load balancing system which tries to abstract from the application specific issues. Thus we have chosen the second approach which is more generic and requires minimal modifications in the application code.

We have developed a meta-algorithm for adaptive load balancing on heterogeneous resources based on benchmarking the available resources capacity (defined as a set of individual resource parameters $\boldsymbol{\mu} = \{\mu_i\}$) and experimental estimation of the application parameter f_c. The algorithm ensures efficient load distribution, thus minimizing the application execution time. The cost function in our case is the experimentally measured execution time, which depends on the distribution of the workload between the participating processors. The target is to experimentally determine the value of f_c that provides the best workload distribution, i.e. minimal runtime of the application mapped to the resources characterized by parameter set $\boldsymbol{\mu}$.

The outline of the load balancing meta-algorithm is as follows:

1. Benchmark the resources dynamically assigned to a parallel application; measure the resource characteristics that constitute the set of resource parameters $\boldsymbol{\mu}$ (available processors power, memory and links bandwidth).
2. Estimate the range of possible values of the application parameter f_c. The minimal value is $f_c^{\min} = 0$, which corresponds to the case when no communications occur between the parallel processes of the application. The maximum value can be calculated based on the following reasoning: For the parallel processing to make sense, that is to ensure that running a parallel program on several processors is faster than sequential execution, the calculation time should obviously exceed communication time. For homogeneous resources this can be expressed as follows:

   $$\frac{T_{comm}}{T_{calc}} < 1 \quad \Leftrightarrow \quad \frac{N_{comm} t_{comm}}{N_{calc} t_{calc}} < 1 \quad \Leftrightarrow \quad f_c^{\max} = 1/\mu$$

   Likewise, for heterogeneous resources the upper limit can be found as:

   $$f_c^{\max} = \max(t^i{}_{calc}) / \min(t^i{}_{comm})$$
3. Run through the range of possible values of f_c with a discrete step. For each value of f_c calculate the corresponding load distribution based on the resource parameters $\boldsymbol{\mu}$ determined in step 1 (details on calculating the load distribution weights will follow this algorithm). With this distribution perform one or a few time steps/iterations, and measure the execution time. Proceed with the next value of f_c for the subsequent iterations, assuring that the simulation continues without delays, with a modified load distribution,
4. Analyze performance results for different values of f_c; find the optimal value f_c^* which provides the best performance of the application (i.e. minimal execution time).
5. Execute further calculations using the discovered f_c^*.
6. In case of dynamic resources where performance is influenced by other factors (which is generally the case on the Grid), a periodic re-estimation of resource parameters $\boldsymbol{\mu}$ and load re-distribution shall be performed.

7. If the application is dynamically changing (for instance due to adaptive meshes, moving interfaces or different combinations of physical processes modeled at different simulation stages) then f_c^* must be periodically re-estimated on the same set of resources.

Periodic re-estimations in steps 6 and 7 shall be performed frequently during the run-time of the application to correct the load imbalance with a reasonably short delay. The minimally required frequency of re-balancing can be estimated by calculating the relative imbalance introduced during the controlled period of time (the number of time steps/iterations).

The combination of $\boldsymbol{\mu}$ and f_c^* determines the distribution of the workload between the processors. To calculate the amount of workload per processor, we assign a weight-factor to each processor according to its processing power, memory and network connection. A similar approach was applied in [21] and in [10] for heterogeneous computer clusters, but the mechanism for adaptive calculation of the weights and application requirements was not developed there. Moreover, the tools developed for cluster systems cannot be used in Grid environments without modifications since static resource benchmarking is not suitable for dynamic Grid resources, where the weights shall be calculated every time the solver is started on a new set of dynamically assigned processors.

Let us assume that for the ith processor: p_i is the available processor performance (e.g. in Flop/s), m_i is the available memory (in MB) and n_i - available network bandwidth to the processor (in MB/s). An individual resource parameter μ_i then can be represented using the values of p_i, m_i, n_i. In a simple case when memory is considered only a constraining factor (and not driving the load balancing process) it is $\mu_i = p_i / n_i$. This resource parameter is widely used in scientific applications where the most important factor is the ratio of the computational power to the network bandwidth. In a more general case, two parameters shall be considered, μ_i and m_i. For the memory-driven applications, the ratio of the available memory to the network capacity of that processor m_i / n_i should play the major role in resource evaluation.

To reflect the processor capacity, we introduce a weighting factor w_i for each processor. It determines the final workload for a processor given by:

$W_i = w_i W$, where W is the total workload.

To determine the weighting factors we introduce parameters c_p, c_m and c_n that reflect computational, memory and communication requirements of the application. Then the weight of each processor is estimated using the following expression:

$$w_i = c_p p_i + c_m m_i + c_n n_i; \quad \sum_i w_i = 1. \tag{5.2}$$

This weighting factor w_i reflects a relative capacity of the resources according to the estimated infrastructure parameter $\mu_i = \mu(p_i, m_i, n_i)$ and the application parameter f_c. The infrastructure parameters μ_i can be determined by a set of benchmark runs before the actual calculations start (but after the resources have been assigned to the application).

Searching through f_c with fixed values of μ_i gives us the optimal value f_c^* which corresponds to the optimal mapping of the workload to the resources.

The parameters c_p, c_m and c_n depend not only on the application characteristics but also on the heterogeneity of the resources. Next we analyze how these parameters and weighting factors w_i are related to f_c and μ_i. Consider a situation where memory is only a constraining factor ($c_m = 0$). Then parameters c_p and c_n shall be proportional to the amount of application communications (computations) and the heterogeneity factors:

$$c_p \sim N_{calc}\varphi_{proc}; \quad c_n \sim N_{comm}\varphi_{net}. \tag{5.3}$$

Here φ_{proc} and φ_{net} are heterogeneity metrics of processors and network links. In case of equal network links the weighting should be done only according to the processors capacity, therefore the network heterogeneity parameter is nullified: φ_{net}=0. Analogously, for homogeneous processors φ_{proc}=0. The heterogeneity metrics of the network and computing resources can be defined as follows:

$$\varphi_{net} = \frac{\sum_{i=1}^{N}(n_i - n_{avg})^2}{N n_{avg}^{\ 2}}, \ \varphi_{proc} = \frac{\sum_{i=1}^{N}(p_i - p_{avg})^2}{N p_{avg}^{\ 2}}. \tag{5.4}$$

Substituting expressions (3) for c_p and c_n in eq. (2), the weights can be re-written as:

$$w_i \sim N_{calc}\varphi_{proc}p_i + N_{comm}\varphi_{net}n_i \tag{5.5}$$

For the trivial cases:

$\varphi_{net} = 0$ (the network is homogeneous): $w_i \sim N_{calc}\varphi_{proc}p_i \sim p_i$

$\varphi_{proc} = 0$ (the processors are homogeneous): $w_i \sim N_{comm}\varphi_{net}n_i \sim n_i$

otherwise

$$w_i \sim N_{calc}\varphi_{proc}p_i + N_{comm}\varphi_{net}n_i \sim p_i + n_i f_c \frac{\varphi_{net}}{\varphi_{proc}}; \quad \sum_i w_i = 1 \tag{5.6}$$

Defining $\varphi = \varphi_{net}/\varphi_{proc}$ as an aggregate heterogeneity metric of resources and keeping in mind that $\mu_i = p_i/n_i$, we get: $w_i \sim p_i(1 + f_c\varphi/\mu_i)$

Considering $\vartheta_i = \frac{\mu_i}{\varphi}$ which combines the characteristics of the resource performance and heterogeneity we get:

$$w_i \sim p_i(1 + f_c/\vartheta_i) \tag{5.7}$$

Knowing the fractional overhead of the application and the heterogeneity level of the resources, we can optimize the workload distribution using this fast weighting technique. To evaluate the efficiency of the workload distribution we introduce the load balancing speedup Θ:

$$\Theta = \frac{T_{non-balanced}}{T_{balanced}} \cdot 100\% , \tag{5.8}$$

where $T_{non-balanced}$ is the execution time of the parallel application without the load balancing, and $T_{balanced}$ is the execution time using load balancing on the same set of resources. This metric is used to estimate the f_c^* that provides the best performance on given resources – the largest value of Θ in a given range of f_c^*. In a non-trivial case we expect to find a maximum of Θ and thus an optimal f_c^* for some workload distribution. Finite and non-zero value of f_c^* means that the application requirements fit best the resources in this particular workload distribution, which minimizes the total run-time of the application. The case of $f_c^* = 0$ while $\varphi \neq 0$ means that the application is totally computation dominated i.e. there is no communication between different processes, and the optimal workload distribution will be proportional only to the computational power of the processors. The case of $\varphi_{net} = 0$ means that we consider the resource infrastructure of heterogeneous processors connected by homogeneous network links and the value of f_c does not play a role – the distribution is again proportional only to the processing power.

In the discussion presented above while deriving eq. (5.4), we considered a simple case where memory requirements only put a Boolean constraint to the allocation of processes on the resources: either there is enough memory to run the application or not. But it can play a role in the load balancing process being one of the determining factors of application performance. This is the case for applications that are able to control memory requirements according to the available resources. In this case there will be additional parameters analogous to f_c and μ_i (or these functions will be more complex), but the idea and the load balancing mechanism remain the same.

5.8. Experimental results of the workload balancing algorithm

To illustrate the approach described in Section 5.7 we present the results obtained for different types of simulation (chemistry-disabled and enabled) of a reactor geometry with 10678 cells on the St. Petersburg Grid site. This site is heterogeneous in both the CPU power and the network connections of the processors (Type IV). There are two 1.8 GHz nodes (nwo1.csa.ru, nwo2.csa.ru) and two dual 450 MHz nodes (crow2.csa.ru, crow3.csa.ru), all having 512 MB RAM. One of the dual nodes (crow3.csa.ru) is placed in a separate network segment with 10 times lower bandwidth (10 Mbit/s against 100 Mbit/s in the main segment). The load balancing tests were performed with a moderate-size problem which does not pose restrictions on required memory, thus the memory influence parameter c_m was reduced to zero and the exploration was done for the application parameter f_c. The link bandwidth between the Master and Slave processors was estimated by measuring the time of MPI_Send transfers of a predefined data block (with the MPI buffer size equal

to 10^6 of MPI_DOUBLEs) during the solver execution, after the resources have been allocated. In these measurements the same logical network topology was used as employed in the solver. The CPU power and available memory were obtained by a function from the *perfsuite* library [22]. To validate the approach presented in the previous section we applied the workload balancing technique for a single simulation running on different sets of heterogeneous resources. The performance was estimated for different possible values of the parameter f_c (hence different weighting and workload distribution). For one simulation type we expect to obtain approximately the same value of the parameter f_c^* (that provides the best performance, see previous section) on different sets of resources. Fig. 5.11 illustrates the load-balancing speedup Θ achieved by applying the workload balancing technique for different values of the parameter f_c on several fixed sets of heterogeneous resources for a light-weighted (chemistry-disabled) simulation.

In Table 5.1 we summarize the combinations of processors dynamically allocated in 4 tests (different sets of resources) and the weights assigned to each processor for the values of f_c^* providing the best execution time, i.e. maximum balancing speedup (see Fig. 5.11, left).

Table 5.1 . Distribution of processors and balancing weights providing the best load balancing speedup for different sets of resources.

Sets of resources	**Weights assigned to each processor:**						**Heterogeneity metrics**		**Balancing speedup**
	nwo1 1.8GHz/ 100Mb/s	**crow2/1** 450MHz/ 100Mb/s	**crow3/1** 450MHz/ 10Mb/s	**crow2/2** 450MHz/ 100Mb/s	**nwo2** 1.8GHz/ 100Mb/s	**crow3/2** 450MHz/ 10Mb/s	φ_{proc}	φ_{net}	Θ
set I 3 processors	0.580	0.274	0.146	-	-	-	0.618	0.606	196 %
set II 4 processors	0.452	0.218	0.112	0.218	-	-	0.638	0.502	182 %
set III 5 processors	0.314	0.146	0.080	0.146	0.314	-	0.591	0.439	201 %
set IV 6 processors	0.278	0.160	0.062	0.160	0.278	0.062	0.618	0.606	207 %

Fig. 5.11 (left) shows that for a given simulation the best performance is delivered by weighting the resources with the value of $f_c \approx$ 0.3-0.4. Noticeably, this corresponds to the value obtained for this simulation during the preliminary analysis on homogeneous resources (compare to results for similar simulations in Section 5.5 Fig. 5.9). The results show that the algorithm gives an increase of balancing speedup Θ up to 207 percent compared to the initial non-balanced version of the code (with homogeneous workload distribution) on the tested resource sets. We can see that the distribution of the workload proportional only to the processor performance ($f_c = 0$) also gives a significant increase of the performance, but introduction of the dependency on application specific communication/computation ratio f_c and resource infrastructure parameters μ_i adds another 40 percent to the balancing speedup Θ.

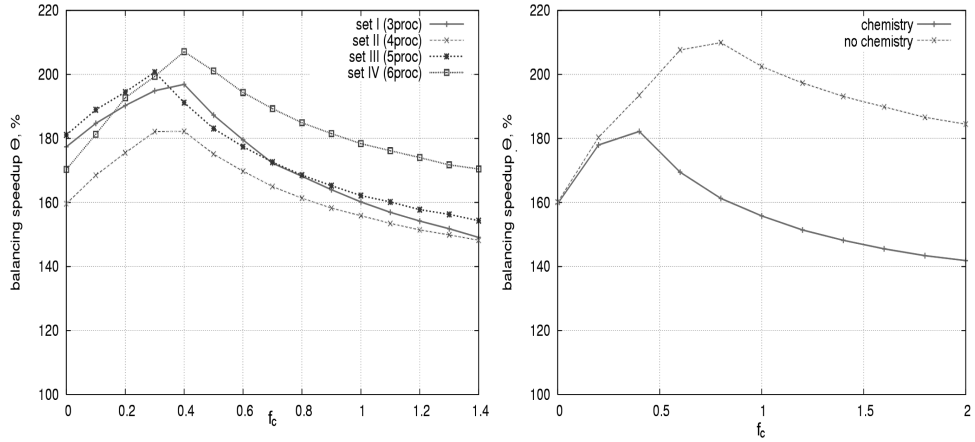


Fig. 5.11. Dependency of the balancing speedup Θ on the resource parameter f_c. Left: single simulation on different sets of resources. Right: different types of simulation on the same set of resources.

Fig. 5.11 (right) shows the dependency of the balancing speedup Θ for different types of simulation (chemistry enabled or disabled) on the same set of resources (set III from Table 5.1). The chemistry-disabled simulation has a higher communication/computation ratio (as was shown also in 5.5 Fig. 5.9). This is clearly seen in the experimental results where chemistry-disabled simulation obtains the highest balancing speedup Θ at higher f_c values. Moreover, the gain in the balancing speedup (maximal value of Θ) is higher for the simulation with a larger fraction of communications. These results illustrate that the introduced algorithm for resource adaptive workload balancing can bring a valuable increase in the performance for communication-intensive parallel programs running on heterogeneous resources.

5.9. Discussion and suggestions for generalized automated load balancing

The introduction of the load balancing technique allowed us to increase the efficiency of the parallel solver on heterogeneous resources. The proposed method of successive estimation of resource infrastructure parameters μ_i and further determination of the application specific f_c shows the possibility of automatic load balancing for applications with unknown internal structure (computations and communications).

Analysis of the results achieved with the workload balancing algorithm suggested that the following issues shall be addressed in order to optimize the balancing technique:

1. To measure the inter-process communication rate, we send a fixed amount of data from the Master to each Slave processor. However in some cases the response of the communication channels to the increasing amount of data is not linearly proportional as shown in Fig. 5.8. For the slower networks this tendency is even more pronounced.

This brings us to a conclusion that the amount of data sent to measure the links performance shall be close to the amount really transferred within the solver for every particular mesh size, geometry and solver type. Another option to estimate the inter-processor communication rate is to analyze the iteration data transfer time during the actual execution. However, this requires significant code modifications and might be undesirable.

2. To properly take into account the memory requirements of each particular instance of a parallel solver, similar reasoning shall be applied as for selecting c_p and c_n: the choice of the c_m coefficient, setting the significance (priority) of the memory factor influence on the application performance, must depend on the type of resources assigned, analogous to the c_p.
3. The specialty of the memory factor is that in addition to this resource-dependency it is strongly influenced by the application features. To take into account the memory requirements of a parallel solver, the weighting algorithm must be enriched by a function measuring the memory requirements per processor for each simulation on each set of resources. In case of sufficient memory on allocated processors, the load balancing can be performed taking into account all the factors (CPU, memory and network) where memory factor is a constraint. After this, another check of meeting the memory requirements on each processor must be performed. In the unfavorable case of insufficient memory on some of the processors, they must be disregarded from the parallel computation or replaced by other, better suited processors. This must be done preferably outside the application, on the level of parallel job scheduling and resource allocation. This brings us to the conclusion that ideally a combined technique shall be developed, where the application-centered load balancing approach is coupled with a system-level resource management.

5.10. Conclusions

In this chapter we address the issue of porting distributed problem solving environments to the Grid, using Virtual Reactor as an example of a complex application. One of the most challenging problems we encountered was porting parallel modules from homogeneous cluster environments to heterogeneous resources of the Grid, specifically the issue of keeping up a high parallel efficiency of the computational components. This problem arises for a wide class of parallel programs that employ homogeneous load distribution algorithms. To adapt these applications to heterogeneous Grid resources, we developed a theoretical approach and a generic workload balancing technique that takes into account specific parameters of the Grid resources dynamically assigned to a parallel job, as well as the application requirements. We validated the proposed algorithm by applying this technique to the Virtual Reactor parallel solvers running on the Russian-Dutch Grid testbed. It is worth noting that the load balancing speedup *does* go through a maximum at $f_c = f_c^*$ as shown in Fig. 5.11. This indicates that the load balancing strategy does find an optimum in the complex parameter space of the heterogeneous application/architecture combination. The clear maximum gives an unbiased guide towards automatic load balancing. The developed approach is well suited for either static or

dynamic load balancing, and can be combined with the Grid-level performance prediction models or application-level scheduling systems [23,24]. To further explore this new load balancing approach, we are currently working on the comparison of the theoretically derived optimization parameters for some specific topologies of parallel applications with those predicted by the heuristic algorithm presented in this chapter.

In order to optimize the resource management strategy of mapping the distributed components of the application problem solving environment, we benchmarked the individual components of the Virtual Reactor on a set of diverse Russian-Dutch Grid resources, and extensively studied the behavior of the parallel solvers with various problem types and input data on different resource infrastructures. The results clearly show that even within one solver different trends can exist in the application requirements and parallel efficiency depending on the problem type and computational parameters, therefore distinct resource management and optimization strategies shall be applied and automated procedures for load balancing are needed to successfully solve complex simulation problems on the Grid.

5.11. References

1. V.V. Korkhov; V.V. Krzhizhanovskaya and P.M.A. Sloot: A Grid Based Virtual Reactor: A case study of parallel performance and adaptive load balancing, Journal of Parallel and Distributed Computing, V. 68/5, May 2008, pp 596-608. http://dx.doi.org/10.1016/j.jpdc.2007.08.010
2. proj-openlab-datagrid-public.web.cern.ch, www.nbirn.net, www.fusiongrid.org, www.globus.org/alliance/projects.php, cmcs.ca.sandia.gov, www.us-vo.org
3. The CrossGrid EU Science project: http://www.eu-CrossGrid.org
4. High Performance Simulation on the Grid project: Dutch-Russian NWO-RFBR-047.016.007
5. V.V. Krzhizhanovskaya, P.M.A. Sloot, and Yu. E. Gorbachev. Grid-based Simulation of Industrial Thin-Film Production. Simulation: Transactions of the Society for Modeling and Simulation International, V. 81, No. 1, pp. 77-85, 2005
6. V.V. Krzhizhanovskaya, M.A. Zatevakhin, A.A. Ignatiev, Y.E. Gorbachev, W.J. Goedheer and P.M.A. Sloot. A 3D Virtual Reactor for Simulation of Silicon-Based Film Production. Proceedings of the ASME/JSME PVP Conference. ASME PVP-Vol. 491-2, pp. 59-68, PVP2004-3120 (2004)
7. A. Barak, G. Shai, and R. Wheeler The MOSIX Distributed Operating System, Load Balancing for UNIX, LNCS, V. 672, Springer-Verlag, 1993
8. K.A. Iskra, F. van der Linden, Z.W. Hendrikse, B.J. Overeinder, G.D. van Albada and P.M.A. Sloot: The implementation of Dynamite - an environment for migrating PVM tasks, Operating Systems Review, V. 34, N 3 pp. 40-55. Association for Computing Machinery, Special Interest Group on Operating Systems, July 2000.
9. G. Shao, R. Wolski and F. Berman. Master/Slave Computing on the Grid. Proceedings of Heterogeneous Computing Workshop, pp 3-16, IEEE Computer Society (2000)

10. S. Sinha, M. Parashar. Adaptive Runtime Partitioning of AMR Applications on Heterogeneous Clusters. In Proceedings of 3rd IEEE Intl. Conference on Cluster Computing, pp. 435-442, 2001
11. R. David et al. Source Code Transformations Strategies to Load-Balance Grid Applications. LNCS V. 2536, pp. 82-87, Springer-Verlag, 2002
12. C. Lu, S.-M. Lau. An Adaptive Load Balancing Algorithm for Heterogeneous Distributed Systems with Multiple Task Classes, International Conference on Distributed Computing Systems (ICDCS'96)
13. Z. Lan, V.E. Taylor, G. Bryan. Dynamic Load Balancing of SAMR Applications on Distributed Systems, Proc. of the ACM/IEEE conference on Supercomputing, 2001
14. Y. Zhang, K. Hakozaki, H. Kameda, K. Shimizu. A performance comparison of adaptive and static load balancing in heterogeneous distributed systems, Proc. of the 28th Annual Simulation Symposium, p. 332, 1995.
15. Fox, G.C., Williams, R.D., Messina, P., 1994, Parallel Computing Works!, Morgan Kaufmann Publishers, Inc.
16. V.V. Krzhizhanovskaya, M.A. Zatevakhin, A.A. Ignatiev, Y.E. Gorbachev, P.M.A. Sloot. Distributed Simulation of Silicon-Based Film Growth. Proceedings of the Fourth International Conference on Parallel Processing and Applied Mathematics (PPAM 2001), Lecture Notes in Computer Science, V. 2328, pp. 879-888. Springer-Verlag 2002
17. Nimrod-G: http://www.csse.monash.edu.au/~davida/nimrod/
18. G. Fox, M. Johnson, G. Lyzenga, S. Otto, J. Salmon, and D.Walker. Solving Problems on Concurrent Processors, volume 1, Prentice-Hall, 1988.
19. J.F. de Ronde; A. Schoneveld and P.M.A. Sloot: Load Balancing by Redundant Decomposition and Mapping, Future Generation Computer Systems, vol. 12, nr 5 pp. 391-407, April 1997
20. J.F. de Ronde. Mapping in High performance Computing. A case study on Finite Element Simulation, PhD thesis, University of Amsterdam, 1998
21. J.D. Teresco et al. Resource-Aware Scientific Computation on a Heterogeneous Cluster. Computing in Science & Engineering, V. 7, N 2, pp. 40-50, 2005
22. R. Kufrin. PerfSuite: An Accessible, Open Source Performance Analysis Environment for Linux. 6th International Conference on Linux Clusters. Chapel Hill, NC. (2005)
23. F. Berman, R. Wolski, H. Casanova, W. Cirne H. Dail, M. Faerman, S. Figueira, J. Hayes, G. Obertelli, J. Schopf, G. Shao, S. Smallen, N. Spring, A. Su, D. Zagorodnov. Adaptive Computing on the Grid Using AppLeS. IEEE Trans. on Parallel and Distributed Systems, vol. 14, no. 4(2003) 369—382
24. X.-H. Sun, M. Wu. Grid Harvest Service A System for Long-Term, Application-Level Task Scheduling. Proc. of 2003 IEEE International Parallel and Distributed Processing Symposium (IPDPS 2003)(2003)

Chapter 6. Problem Solving Environment *

6.1. Introduction

In the previous chapters we have introduced step by step, with increasing complexity, various models, implementations and numerical results of processes involved in Plasma Enhanced Chemical Vapor Deposition resulting in what we called a 'Virtual Reactor'. In this chapter we describe an *integrated, flexible* environment to support such virtual reactors over *distributed* computer systems.

There is a clear trend in modeling and simulation away from rigid simulation codes treating a specific fixed aspect, towards more sophisticated flexible problem-solving environments (PSEs) [1] because, as W.D. Hillis (MIT, 1982) puts it: '*Progress in natural science comes from taking things apart, progress in computer science comes from bringing things together*'. These environments are widely considered to be an essential emerging technology with high impact across all fields of science and engineering. PSEs are crucial to successfully address all kinds of multidisciplinary problems, where vast amounts of distributed data need to be managed and processed to discover patterns and knowledge contained within it [2]. Various research groups and scientific software companies subscribe to the relevance of these fully integrated simulators (like the Virtual Reactor) [3].

A PSE provides a complete integrated software environment for composing, compiling, running, controlling and visualizing applications in a specific area. It incorporates many features of an expert system through which it provides extensive assistance to users [4]. Processing, visualization and integration of information from various sources play an increasingly important role in science and engineering. These sources can be widely distributed, and the computational requirements can be highly variable, both in the type of resources required and the processing demands put upon these systems. Grid technology as integrating middleware is a major cornerstone of today's PSEs [5]. By offering a unified means of access to different and distant computational and instrumental resources, it brings unprecedented possibilities and benefits. Connectivity between distant locations, interoperability between different kinds of systems and resources, and high levels of computational performance are some of the most promising characteristics of the Grid. Grid technology provides dedicated support such as strong security, distributed storage capacity, and high throughput over long distance networks. Besides these immediate benefits, the computational resources of the Grid provide the required performance for large-scale simulations and complex visualization and collaborative environments, which are expected to become of major importance to many areas of computational physics. The Globus project [6,7] is perhaps the best-known example of the core software implementation of these Grid functionalities.

* Parts of this chapter were published in [24-27]

The importance of Grid-based PSE research and applications is recognized by international research foundations such as the European Science Foundation (Euresco) through the ESF-PSE initiative, and the NSF initiative on PSEs in the USA. PSEs are becoming *the* key technology to facilitate full exploitation of the unprecedented computational power offered by the Grid revolution [8]. Researchers are investigating many ideas to make this vision reality [9-12].

Recently, lots of efforts have been invested in developing Component Based Environments such as CCA [13], H2O [14], Mocca [15] and ProActive [16]. The concept of common component architectures is a promising approach to build PSE's due to features like light-weighted software components, support for dynamic behavior, scalability, etc. This approach can be beneficially applied for defining interoperable and reusable simulation models that can be automatically arranged into distributed interactive applications [17]. The advantages of these component technologies were used while building the Virtual Reactor PSE that is discussed in this thesis. Integration in to a Grid based PSE was done through the use of CrossGrid middleware [18] and previously developed interactive simulation and visualization tools developed for CrossGrid applications [19-21]. The PSE considered here is generic in that it can handle High Performance Computing (HPC) applications (one parameter setting calculated as fast as possible) as well as High Throughput Computing (HTC) (sweeping various parameter settings) applications and supports interactivity.

Within the CrossGrid team there is a focus on the development of Grid middleware components, tools and applications with a special focus on parallel and interactive computing. Interaction in this context, refers to the presence of a human in a processing loop, and a requirement for near real-time response from the computer system. The CrossGrid testbed largely benefits from the DataGrid [22] experience in testbed setup and Globus [7] middleware distributions.

As a design template, we used a generic architecture for research in e-Science, see [23], where 'information systems integrate available data with data from specialized instruments into distributed repositories and computational models are executed using the integrated data, providing large quantities of model output data, which is mined and processed in order to extract useful knowledge'.

The next sections describe the Virtual Reactor architecture, its most relevant components, the functionalities and the implementation into the Grid software.

6.2. Virtual Reactor Architecture

6.2.1. General description of the *application* components

The Virtual Reactor PSE includes the basic *application* components (modules) for the reactor geometry design; computational mesh generation; plasma, flow and chemistry simulation; editors of chemical processes and gas properties connected to the corresponding databases; pre- and postprocessors, visualization and data archiving modules. In addition to the *application* components we have the Grid *middleware* 'components' to support distributed simulations on the Grid. From the perspective of the Virtual Reactor PSE, the

Grid middleware can be considered as a set of 'components' that provide transparent access to the Grid resources and services.

In order to make the system user-friendly we need to hide the complexity of the underlying components. For that we have developed an advanced graphical user interface (GUI) that seamlessly integrates the disparate modules into *one* transparent user environment, which is presented to the user via a Web-interface and a Grid portal [24]. The architecture supports interaction on various levels: interaction with the workflow through the user interface, interactive visualization, control over the simulation processes and interactive job control on the middleware level.

The Virtual Reactor *application* components and their interdependencies are sketched in Fig. 6.1. It's important to note that there are in fact 5 solvers that were described in the previous chapters (1D and 2D plasma, 1D, 2D and 3D flow) that can be used in different combinations depending on the level of detail required.

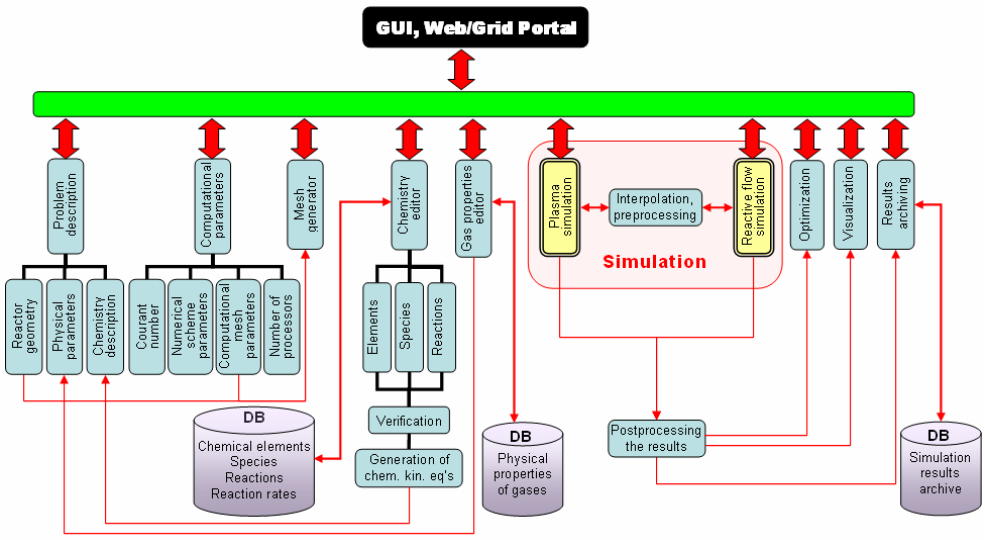


Fig. 6.1. Virtual Reactor *application* components and their interdependencies

The arrows in Fig. 6.1 indicate some of the interdependencies of the components. For instance, the solvers for plasma and reactive flow simulations use as input data:

- the description of the physical and chemical properties created by the editor components;
- parameters of the PECVD processes (like sticking probabilities of the species);
- the generated computational mesh;
- computational parameters (e.g. Courant number, numerical scheme parameters, number of processors for parallel computing, etc.).

The results of the simulations can be stored via an archiving component and retrieved later for comparison with new computational experiments. All the components are

stand-alone modules that can be used separately or in various arrangements in other studies. They are platform-independent and can be compiled and run on any operating system using public-domain compilers and libraries. Some of the most relevant components will be described below.

6.2.2. User interface components

The GUI allows setting up a workflow that describes the problem to be simulated, sets the reactor and computational parameters, controls the simulation process, visualizes and analyzes the simulation results, and gives access to the databases and result archives. An advanced visualization system provides a graphical representation of the results in real-time or postponed mode on different visualization systems. The GUI was implemented with the use of C/C++ and a platform independent GTK+ graphic library [28] as well as the Glade user interface builder [29].

To provide remote access to the Virtual Reactor, a Web interface was developed, using standard client-server technology. For this interface, we created HTML pages, Java applets, JavaScripts and CGI scripts. This system, like a local GUI, provides full control over the simulations. Fig. 6.2 shows some screenshots of the user interfaces described above.

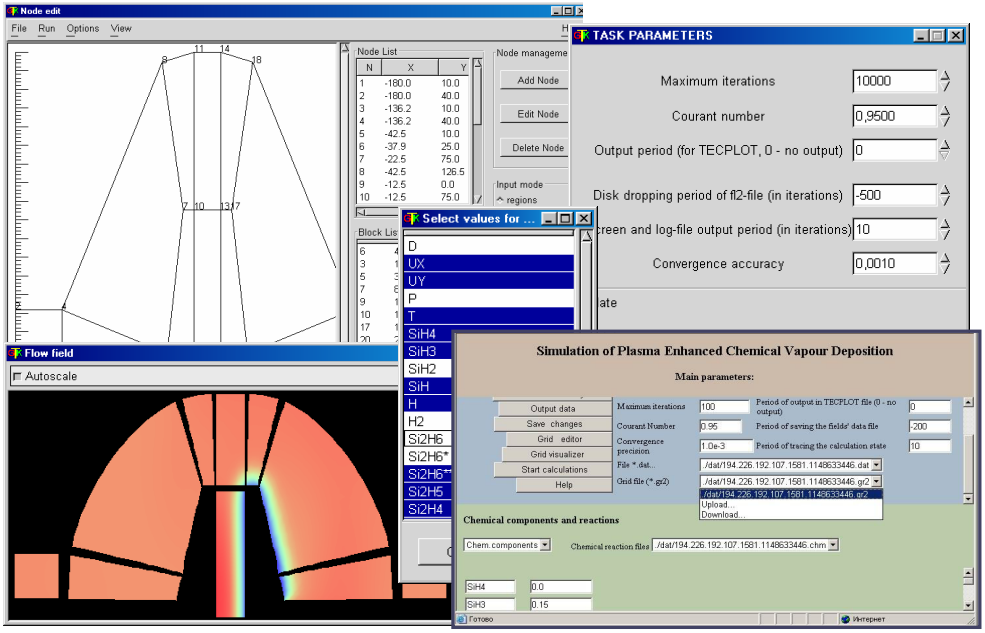


Fig. 6.2. Graphical user interface, stand-alone and Web-based versions.

6.2.3. Chemistry and gas editors components

The Chemistry Editor component provides access to the database of chemical elements, species and reactions, and allows selecting a set of chemical processes for building a customized chemistry model. The verification system incorporated into the module checks the balance of chemical elements in each new reaction introduced by the user, verifies the correctness of the reaction set and removes chemical species (and correspondent reactions) which have a zero net balance. Finally, the module generates a set of differential equations describing the chemical kinetics based on the system of reactions selected by the user from the data base.

The Gas Properties Editor provides access to the database of physical properties of gases and facilitates the introduction of new species into the user chemistry model. One of the important features of this Gas Editor is that it can suggest (based on the available kinetic data) approximate values of the specific heat of formation, the dependency of the specific heat capacity on temperature, and the Lennard–Jones potential parameters for unknown species of higher silanes (Si_nH_m, with n,m > 5). Fig. 6.3 shows some screenshots of these two editors.

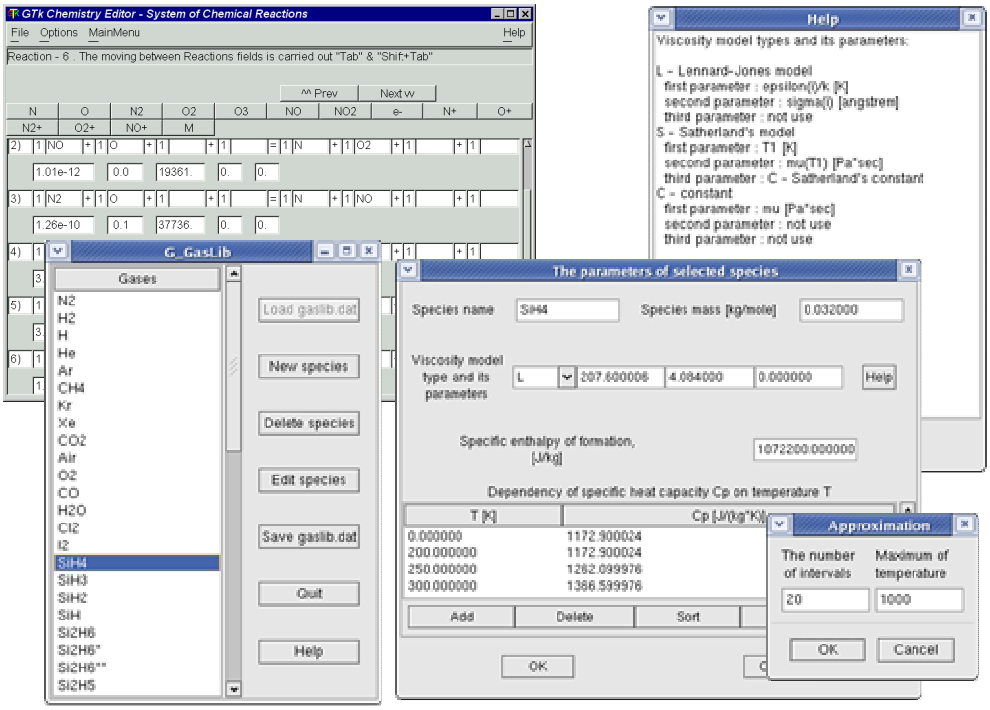


Fig. 6.3. Chemistry and Gas Properties Editors

6.2.4. Visualization components

Different visualization systems were studied and developed for visual analysis of the results. First a commercial package Amtec Tecplot [30] was used. Although it satisfied

most of the basic needs for results visualization, it lacks the flexibility to adjust and incorporate it into some other environment. To overcome this, a Java-based multi-purpose visualizer [24] was developed and incorporated into the Virtual Reactor. This component performs all essential functions of visualizing 2D and 3D objects and variable scalar and vector fields (e.g. density, pressure, temperature, species mass fractions, velocity vectors, etc.). A distinctive feature of the visualizer is that it is specially adapted to the multi-block computational meshes used for complex geometry simulations. For time-dependent simulations, the visualizer provides an option of creating animations. A screenshot of this visualizer is shown in Fig. 6.4.

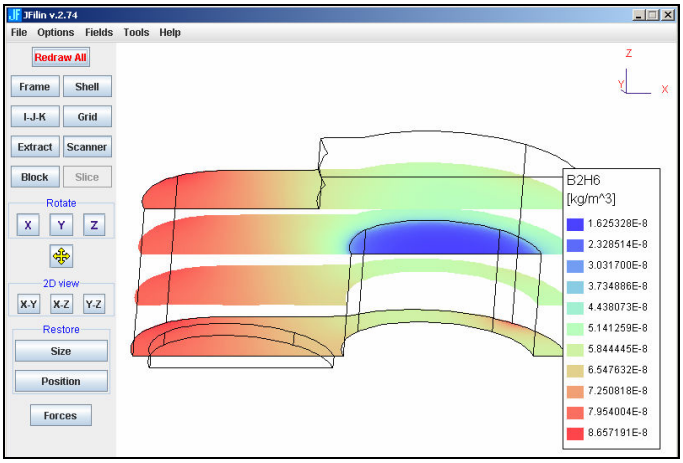


Fig. 6.4. Java visualizer

Another visualization component supports advanced 3D Virtual Reality for a wide range of visualization systems: a personal computer, the virtual reality DRIVE, the 3D immersive environment CAVE and the Personal Space Station, without the actual need to

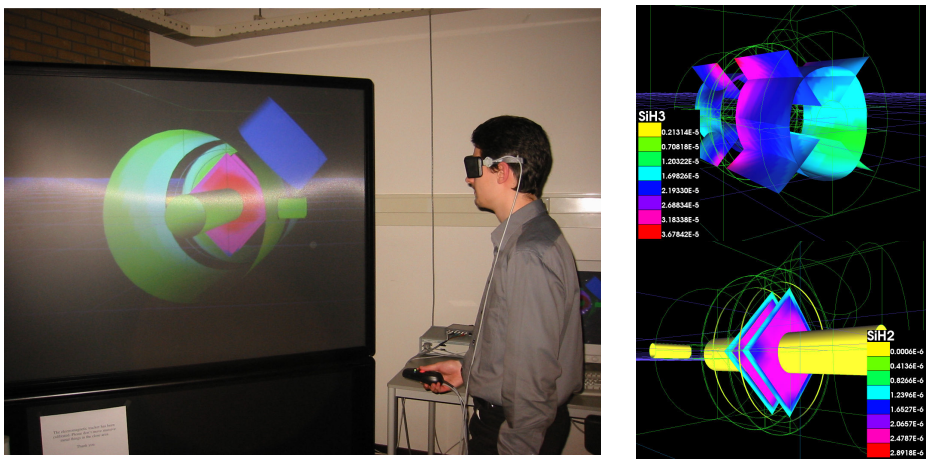


Fig. 6.5. Virtual reality interaction on the DRIVE visualization front-end of the Virtual Reactor

change the code [31,32]. The 3D immersive visualization system was developed based on the VTK and the CaveLib libraries. The DRIVE System uses Immersive Projection Technology with one large screen, active stereo with shutter glasses, electromagnetic tracking and multi-modal interaction devices. A snapshot of visualization on the DRIVE is shown in Fig. 6.5.

6.2.5. Grid middleware components

The Virtual Reactor PSE integrates the *application* components described above with the previously developed interactive simulation environment [32,33], using the common component technology [9] and Grid middleware. This is schematically shown in Fig. 6.6. The Grid middleware is based on the CrossGrid and the DataGrid testbeds [18,22] which use Globus distributions. The CrossGrid testbed shares resources across sixteen European sites [18]. The sites range from relatively small computing facilities in universities to large computing centers. National research networks and the multi-gigabit pan-European network Geant [34], assure interconnectivity between all sites. The actual network used for the Virtual Reactor, includes a local step inside the university via Fast or Gigabit Ethernet, a jump via a national network provider at speeds that range from 34 Mbit/s to 622 Mbit/s or even Gigabit to the national node, and a link to the Geant network at 155 Mbit/s to 2.5 Gbit/s.

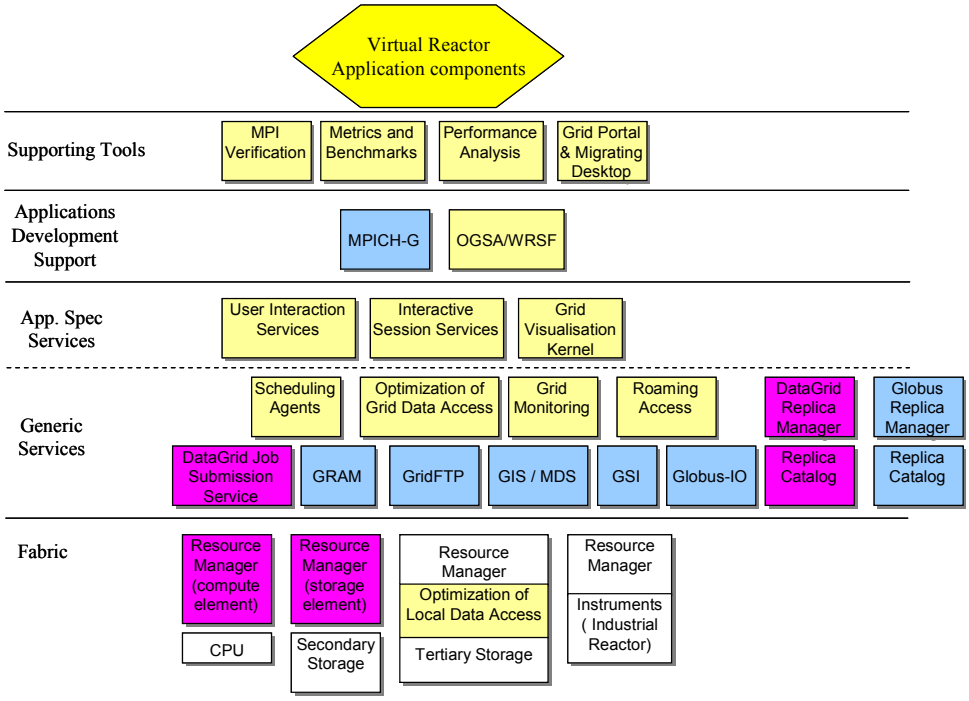


Fig. 6.6. The architecture of the Virtual Reactor *application* and Grid middleware components. The top hexagon indicates all the *application* components described in the subsections above. (1) Yellow: CrossGrid components, (2) Blue: Globus components (3) Red: DataGrid components. For more detailed description of this architecture see [23] and references in there.

A key component of the used Grid architecture is the Grid portal and the Migrating Desktop (MD) as its front-end [18]. The MD was implemented as a Web-based (programmed with Java and JavaScript) user interface for application management, grid and job monitoring, data and metadata management. It includes an authentication mechanism and advanced grid tools. It provides a transparent user work environment, independent of the operating system and hardware. It allows the user to access the Virtual Reactor *application* components, Grid and local resources from remote computers, to run applications, manage data files, and store personal settings, independent of the location or the terminal type. In addition, there is a set of performance analysis components such as OCM-G and G-PM [18] that can be used to monitor and fine-tune the performance of the parallel distributed simulations.

In this architecture, program executions are performed using the Globus job submission capabilities. The parallel solvers use MPICH-G2 as an implementation of the MPI for Grid [35], Globus I/O for inter-process communication and Globus DUROC for resource co-allocation. The Resource Manager in Fig. 6.6 is a component of the DataGrid that has been modified by the CrossGrid project.

This Grid middleware supports different levels of interactivity [36]:

- Interactivity with the Client/Grid System: a user can continuously interact with a Grid client (MD, portal, application GUI, editors, other dedicated clients) without waiting for the conclusion of the jobs already submitted. This interactivity does not need any specialized infrastructure from a Grid testbed point of view.
- One-Way Interactivity with a running application: a user can see the output of the application running on the Grid via the MD client/GUI, synchronously with the application. This interactivity is assured through the infrastructure deployed on the Grid.
- Two-Way Interactivity with the running application: a user via the MD and application GUI can steer the running simulation, either providing some input data on-line as requested by the application or asynchronously, suspending the simulation, changing some input data and resuming it. At the same time, application output data is forwarded to the MD client. This interactivity was implemented using the Condor ByPass (Job Shadow) mechanism [37].

We have incorporated the Virtual Reactor into the Grid using the above described system. We achieved secure Grid access, resource discovery and registration, Grid data transfer, application initialization, editing physical and chemical properties DBs, parameter specification, job submission, distributed simulations and advanced 3D visualization.

We have tested the complete problem-solving environment on a number of tightly coupled clusters as well as on distributed computer systems (the CrossGrid and Russian-Dutch Grid testbeds), where separate modules of the PSE (databases, archives, computational core, visualization kernel and the user interface) were located and operating at different sites.

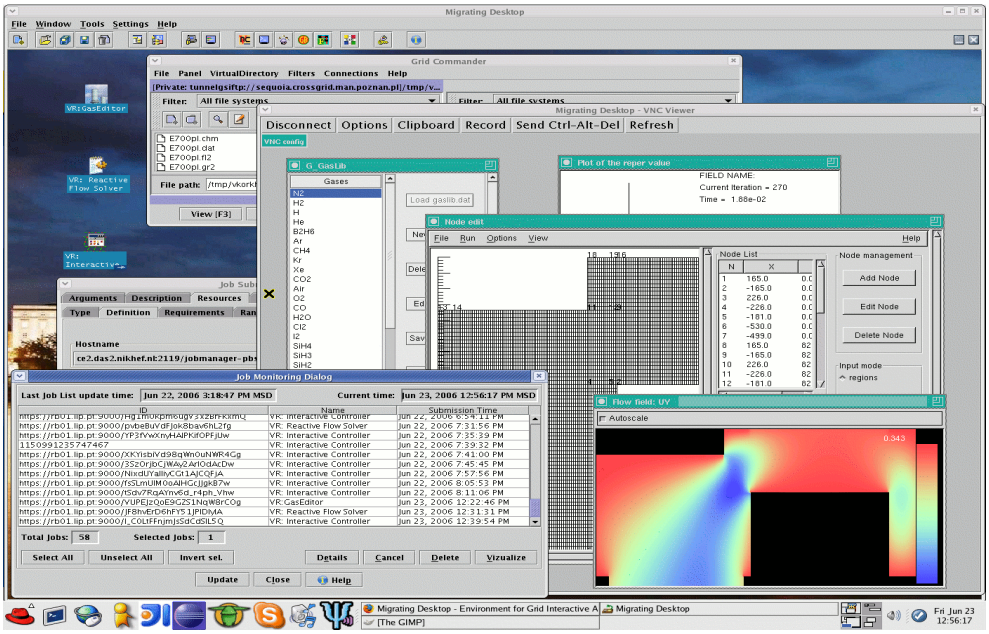


Fig. 6.7. Virtual Reactor on the Grid: Migrating Desktop Grid portal

6.3. Conclusions

The Virtual Reactor was implemented on the Grid and tested against different testbed configurations. We achieved secure Grid access, resource discovery and registration, Grid data transfer, application initialization, editing physical and chemical properties DBs, parameter specification, job submission, distributed simulations and advanced 3D visualization. The foremost conclusion of the presented work is that the core parallel solver can be efficiently exploited on clusters with high-speed interconnects within the RDG environment. The other components of the Virtual Reactor (loosely coupled or fully decoupled) can be distributed across the other Grid resources, thus maximizing the overall efficiency of the whole application.

In addition to the High Performance Computing considerations discussed here, there are at least two more ‘performance’ aspects that play a role in such generic PSE’s as the Virtual Reactor. One is that we might want to run a large set of simulations with slightly different parameters. Here systems like Nimrod and Condor can play an important role. Another ‘performance’ qualifier is given by the relative ease of setting up a computer simulation experiment in the first place as well as the repetition of such experiments. This requires intuitive methods to dynamically streamline the underpinning processes depending on their availability, their reliability, and the specific interests of chemists, engineers, researchers, and other end users. Scientific workflows, in which a workflow language expresses the flow of data and action from one step to another, provide one option for capturing such methods (see for instance [38,39]). Recently these types of experimental environments have drawn lots of attention. There are even efforts on the way to combine

parameter sweep systems (like Nimrod) with advanced workflow systems like Kepler or the VLe [Virtual Laboratory for e-Science Bsik project [40], see for instance [41]. Although these emerging technologies are very promising, further implementation of it in the PSE described here is out of scope of this thesis and can be considered as future work.

6.4. References

1. http://www.esf.org/esf_euresco_conference.php?language=0&domain=1&conference=139&meeting=3
2. P.M.A. Sloot; D. Frenkel; H.A. Van der Vorst; A. van Kampen; H.E. Bal; P. Klint; R.M.M. Mattheij; J. van Wijk; J. Schaye; H.-J. Langevelde; R.H. Bisseling; B. Smit; E. Valenteyn; H.J. Sips; J.B.T.M. Roerdink and K.G. Langedoen: "Computational e-Science: Studying complex systems in silico". A National Coordinated Initiative. White Paper. February 2007.
3. http://www.cfdrc.com/applications/semicon/plasma.html , http://www.fluent.com/about/news/pr/pr10.htm , http://www.semitech.us/consulting/pvt_growth/models/ , http://www.softimpact.ru/VIROOS-engl.html
4. Houstis, E., Gallopoulos, E., Bramley, R., Rice, J., 1997, "Problem-Solving Environments for Computational Science," IEEE Computers in Science and Engineering, 4, (3) pp. 18-21.
5. Fox, G., 2003, "Grid Computing environments," IEEE Computers in Science and Engineering, 10, pp. 68-72.
6. Globus: http://www.globus.org
7. Foster, I., Kesselman, C., 1997, "Globus: A Metacomputing Infrastructure Toolkit," Intl J. Supercomp Appl, 11(2), pp.115-128.
8. Houstis, E.N. and. Rice, J.R, 2000, "Future Problem-solving Environments for Computational Science," Math. Comp. Sim. 54, pp 243-257.
9. Marinescu, D.C. and Bölöni, L., 2000, "A component-based architecture for problem-solving environments," Math. Comp. Sim. 54, pp. 279-293.
10. Walker, D.W., Li, M., Rana, F., Shields, M.S., and Huang, Y., 2000. "The software architecture of a distributed problem-solving environment," Concurrency: Pract. Exp., 12, pp. 1455-1480.
11. Johnson, C., Parker, S.G., Hansen, C., Kindlmannn, G.L., Livnat, Y., 1999, "Interactive Simulation and Visualization," IEEE Computer, December, pp. 59-65.
12. http://www.teragrid.org/
13. The Common Component Architecture Forum, 2004. http://www.cca-forum.org/
14. D. Kurzyniec, T. Wrzosek, D. Drzewiecki, and V. S. Sunderam. Towards Self-Organizing Distributed Computing Frameworks: The H2O Approach. Parallel Processing Letters, 13(2):273-290, 2003.
15. ProActive project homepage. http://www-sop.inria.fr/oasis/ProActive

16. Maciej Malawski, Dawid Kurzyniec, Vaidy S. Sunderam: MOCCA - Towards a Distributed CCA Framework for Metacomputing. IPDPS 2005
17. Bubak, M., Funika, W., Balis, B., Gubala, T., Malawski, M., Radecki, M., Rycerz, K., Smetek, M., Szepieniec, T., CYFRONET Contribution to CoreGRID: Problem Solving Environments, Tools and GRID Systems, in: Bubak, M., Turala, M., Wiatr, K. (Eds.), Proceedings of Cracow Grid Workshop - CGW'04, December 13-15 2004, ACC-Cyfronet UST, 2005, Krakow, pp. 45-57.
18. CrossGrid EU Science project: http://www.eu-CrossGrid.org/
19. P.M.A. Sloot; A. Tirado-Ramos; A.G. Hoekstra and M. Bubak: *An Interactive Grid Environment for Non-Invasive Vascular Reconstruction*, in 2nd International Workshop on Biomedical Computations on the Grid (BioGrid'04), in conjunction with Fourth IEEE/ACM International Symposium on Cluster Computing and the Grid (CCGrid2004), (CD-ROM IEEE Catalog # 04EX836C) IEEE, Chicago, Illinois, USA, April 2004. ISBN 0-7803-8431-8.
20. A. Tirado-Ramos; P.M.A. Sloot; A.G. Hoekstra and M. Bubak: *An Integrative Approach to High-Performance Biomedical Problem Solving Environments on the Grid*, Parallel Computing, (special issue on High-Performance Parallel Bio-computing) vol. 30, nr 9-10 pp. 1037-1055. (Chun-Hsi Huang and Sanguthevar Rajasekaran, editors), 2004.
21. Iskra K.A., Belleman R.G., Van Albada G.D., Santoso J., Sloot P.M.A., Bal H.E., Spoelder H.J.W. and Bubak M., 2002, "The Polder Computing Environment, a system for interactive distributed simulation," Concurrency and Computation: Practice and Experience (Special Issue on Grid Computing Environments), 14, pp. 1313-1335.
22. Hoschek, W., Jaen-Martinez, J., Samar, A., Stockinger, H., Stockinger, K., 2000, "Data Management in an International DataGrid Project," Proc. of the IEEE/ACM International Workshop on Grid Computing, Dec. 17-20, 2000, Bangalore, India.
23. P.M.A. Sloot; A. Tirado-Ramos; I. Altintas; M.T. Bubak and C.A. Boucher: *From Molecule to Man: Decision Support in Individualized E-Health*, IEEE Computer, (Cover feature) V. 39, N 11 pp. 40-46. November 2006.
24. V. V. Krzhizhanovskaya, P.M.A. Sloot and Yu. E. Gorbachev. *Grid-based Simulation of Industrial Thin-Film Production*. Simulation: Transactions of the Society for Modeling and Simulation International, Special Issue on Applications of Parallel and Distributed Simulation in Industry, January 2005, V. 81, No. 1, pp. 77 – 85
25. Krzhizhanovskaya V. V., Gorbachev Yu. E. and Sloot P.M.A. A Grid-Based Problem-Solving Environment for Simulation of Plasma Enhanced Chemical Vapor Deposition. Proceedings of the International Conference "Distributed Computing and Grid Technologies in Science and Education". Publ: JINR, Dubna, 2004, pp. 262-271
26. Gorbachev Yu.E., Zatevakhin M.A., Ignatiev A.A., Krzhizhanovskaya. Virtual Laboratory for Research and Education. XI International Conference on High Technologies and Quality of Education and Science, St. Petersburg Polytechnic University. 27-28 February 2004. Publ.: SPbSPU, St. Petersburg, 2004, pp. 59-60

27. V.V. Korkhov; V.V. Krzhizhanovskaya and P.M.A. Sloot: A Grid Based Virtual Reactor: A case study of parallel performance and adaptive load balancing. Journal of Parallel and Distributed Computing, V. 68/5, May 2008, pp 596-608. http://dx.doi.org/10.1016/j.jpdc.2007.08.010
28. http://www.gtk.org/
29. http://glade.gnome.org/
30. http://www.tecplot.com/
31. http://www.science.uva.nl/research/scs/projects/visualisation/index.php
32. R.G. Belleman: *Interactive Exploration in Virtual Environments*, PhD thesis, University of Amsterdam, Amsterdam, The Netherlands, April 2003.
33. Belleman, R.G. and Sloot, P.M.A., 2000, "The Design of Dynamic Exploration Environments for Computational Steering Simulations," Proc. of the SGI Users' Conference, pp. 57-74.
34. Geant: http://www.dante.net/server.php?show=nav.007
35. Karonis, N., Toonen, B., and Foster, I., 2003, "MPICH-G2: A Grid-Enabled Implementation` of the Message Passing Interface," Journal of Parallel and Distributed Computing.
36. Personal communication with Stefano Beco – Marco Sottilaro, Datamat S.p.A.
37. Douglas Thain and Miron Livny, "Multiple Bypass: Interposition Agents for Distributed Computing", Journal of Cluster Computing, volume 4, pages 39-47, 2001.
38. L. Altintas et al., “A Framework for the Design and Reuse of Grid Workflows,” *Proc. Scientific Applications of Grid Computing*, (SAG 04), LNCS 3458, Springer, 2005, pp. 119-132.
39. F. Neubauer, A. Hoheisel, and J. Geiler, “Workflow-Based Grid Applications,” *Future Generation Computer Systems*, Jan. 2006, pp. 6-15.
40. The Virtual Laboratory for e-Science project: http://www.vl-e.nl
41. AppLeS Parameter Sweep Template (APST) Project: http://grail.sdsc.edu/projects/apst/

Chapter 7. Summary and Conclusions

The work presented in this thesis describes the design and development of a computational modeling and simulation environment to study Plasma Enhanced Chemical Vapor Deposition (PECVD). We call this environment the 'Virtual Reactor'. The PECVD processes are notoriously complicated with many intricate multi-scale, multi-physics and multi-chemical aspects. The Virtual Reactor should support a better understanding of these processes and assist in designing and optimizing real experiments. The challenge was to find the right level of detail and abstraction of all the related processes that still gives us an adequate, reliable and computationally efficient view into the PECVD processes.

Chapters 2 and 3 describe the physics and chemistry as well as the numerical models needed to study PECVD with flow in 1, 2 and 3 dimensions and with a 1-dimensional plasma model. In our simulations, we studied the parameters considered to be important for either adjusting or assessing reactor performance, such as the flow rate, reactor volume, silane dilution, discharge gap and substrate separation in plasma configuration. Comparison with experimental data demonstrated the ability of the model to predict the film growth rate and concentrations of individual components with reasonable accuracy. It was shown that the widely used technique of diluting silane with molecular hydrogen increases the growth rate and reduces production of higher silanes, thus making this technology more economical and ecologically clean. Numerical simulation indicated that the effective decomposition of silylyl outside the discharge zone reduces its contribution to film growth as the substrate is moved away from the discharge. The analytical expressions obtained for silyl and silylyl fluxes and for the profile of atomic hydrogen closely approximate the results of our numerical simulations. Real-life reactors are often not symmetrical, therefore the influence of reactor geometry on the actual gas flow needed further study. For that we developed a 2D and 3D numerical gas flow simulation, to study in detail the influence of reactor configuration on the flow field and on the deposition homogeneity. We observed that in some process conditions both 3D and 2D simulations predict that radial distribution of main parameters (e.g. species concentrations) along the wafer is almost ideally symmetrical with the axis of symmetry in the center of the wafer. This indicates that the homogeneity of these parameters near the wafer was mainly defined by the diffusive transport and not the convection. In some other cases, however (with a high flow rate) the symmetry was broken, which means that a full 3D simulation is essential to correctly predict the homogeneity of the deposition process. The Virtual Reactor allowed us to study also the film degradation processes and to find the featuring trends in system response to the variation of physical and chemical parameters. We were able to indicate the mechanisms responsible for the layer composition at the finishing stage of the film production cycle.

We did however observe a mismatch between the experimental and simulated distribution of the film thickness along the substrate. Additional numerical experiments indicated that the correct (experimental) shape can be recovered by adjusting the electron density distribution rather than gas flow parameters and substrate temperature distribution.

We therefore conclude that a 2D discharge model is needed. This model we implemented in Chapter 4. The simulation results in this Chapter show a qualitatively good agreement of the simulated results with the experimental data in the film distribution shape and the observed thickness variation. We also obtained a good correlation of measured film properties with the simulated ion and radical fluxes towards the substrate. In addition we studied the influence of pressure, temperature and plasma discharge parameters (power and frequency) on the PECVD processes in a wide range of parameters. These results clearly indicated that the Virtual Reactor is capable of predicting trends in the most relevant film growth parameters and can be used for studying, predicting and optimizing the PECVD technology. In addition it was shown that even small changes in the inlet and outlet position can significantly affect the uniformity and composition of the growing film. We conclude from this that the deposition characteristics are sensitive to the *complete* experimental parameter space.

After having established reliable numerical models for the PECVD process we turned our attention to the performance and usability aspects of the Virtual Reactor (Chapters 5 and 6). Here the design considerations were that the simulations should run efficiently both in the *capacity* and in the *capability* mode of the Virtual Reactor. We therefore studied in depth the computational complexity of the PECVD simulations and developed novel parallelization methods to speed up the numerical processes. In this optimization approach we take into account the execution of the Virtual Reactor on tightly coupled (parallel clusters) and loosely coupled (grid-based) computer systems. One of the most challenging problems was porting the parallel modules from homogeneous cluster environments to the heterogeneous resources of the Grid, whilst keeping up the parallel efficiency of the computational components. To solve the load-balancing problem, we developed a theoretical framework and a generic workload balancing technique that takes into account specific parameters of the Grid resources dynamically assigned to a parallel job, as well as the application requirements. We validated the proposed algorithm by applying this technique to the Virtual Reactor parallel solvers running on the Russian-Dutch Grid test-bed. In order to optimize the resource management strategy of mapping the distributed components of the application problem solving environment, we benchmarked the individual components of the Virtual Reactor on a set of diverse Russian-Dutch Grid resources, and extensively studied the behavior of the parallel solvers with various problem types and input data on different resource infrastructures. The results show that the method developed supports dynamic adaptation of the workload over the dynamically changing resources.

Finally, we integrated all the numerical models and computational methods into one Problem Solving Environment. The Virtual Reactor was implemented on the Grid using the Cross-Grid infrastructure and tested against different test-bed configurations. We achieved secure Grid access, resource discovery and registration, Grid data transfer, application initialization, editing physical and chemical properties databases, parameter specification, job submission, distributed simulations and advanced 3D visualization. The foremost conclusion is that the core parallel solver can be efficiently exploited on clusters with high-speed interconnects. The other components of the Virtual Reactor (loosely coupled or fully decoupled) can be distributed across the other Grid resources, thus maximizing the overall

efficiency of the whole application. To increase the efficiency of the solver on heterogeneous resources, the workload balancing techniques developed in the previous chapter was successfully incorporated into the PSE, thus supporting the adaptation of data distribution among the resources according to the network connectivity of the nodes and to the available processing power.

Starting from simplified models, via numerical methods and computer codes, this work has grown into an advanced simulation environment, which has been extensively used in a number of international projects. Recently this work branched into three focused projects aiming at further development of the models, the algorithms and the grid-based environment. Lots of work is planned, and as the history of science shows, solving one problem opens the door to a new one, ever more challenging and exciting..

Chapter 8. Nederlandse Samenvatting

Dit proefschrift beschrijft het ontwerp en de ontwikkeling van een modelleer- en simuleeromgeving waarmee complexe processen uit Plasma Enhanced Chemical Vapor Deposition (PECVD) bestudeerd kunnen worden. Het resulterende computer simulatiesysteem noemen we de 'Virtuele Reactor'. PECVD processen zijn notoire gecompliceerde processen met vele multi-schaal (ruimte en tijd), multi-fysische en multi-chemische aspecten. De Virtuele Reactor beoogd tot een beter begrip van deze processen te leiden en te helpen in de ontwikkeling en optimalisatie van de werkelijke (niet gesimuleerde) experimenten. De uitdaging lag er voornamelijk in om het juiste nivo van abstractie te vinden voor al de relevante processen, zodanig dat de simulaties ons een adequate, betrouwbare en computationeel efficiënte kijk geven op de PECVD processen.

Hoofdstukken 2 en 3 beschrijven de fysica en de chemie alsook de numerieke modellen welke essentieel zijn om stroming in 1, 2 en 3 dimensies met een één dimensionaal plasma model in PECVD te bestuderen. In onze simulaties onderzoeken we welke parameters van belang zijn voor de reactor performance. Hierbij kijken we ondermeer naar gas stromingssnelheid, reactor volume, silaan concentratie, ontladingsruimte en substraat scheiding. Vergelijking met experimentele data laat zien dat het model in staat is de film-groei snelheid en de concentraties van de individuele componenten met redelijke nauwkeurigheid te voorspellen. De simulaties lieten verder zien dat de veel gebruikte techniek om silaan te verdunnen met moleculair waterstof de groeisnelheid doet toenemen en de productie van hogere silanen doet afnemen. Dit is een opmaat voor een schonere en een meer economisch en ecologische verantwoord productie proces. Numerieke simulaties geven aan dat de effectieve decompositie van silylyl buiten de ontladingszone de bijdrage aan het groeiproces reduceert indien het substraat verder weg komt van de feitelijke ontlading. We vonden een goede overeenkomst tussen de afgeleide analytische vergelijkingen en de numerieke simulaties welke we verkregen voor de silyl en silylyl flux en voor het atomair waterstof profiel.

Omdat reactoren vaak niet symmetrisch zijn besloten we de invloed van de reactorgeometrie op de gasstroming verder te bestuderen. Hiertoe ontwikkelden we een 2D en 3D gasstroming computersimulatie, waarmee we de invloed van de reactorconfiguratie op het stromingsveld en op de homogeniteit van de depositie bestudeerden. Deze simulaties lieten zien dat onder sommige condities de 2D en 3D simulaties een bijna volkomen symmetrische radiale verdeling van de concentraties van de chemische componenten over de wafer voorspellen. Hieruit concluderen we dat de homogeniteit van deze componenten in de buurt van de wafer voornamelijk bepaald wordt door transport middels diffusie en niet door convectie. Echter, in enkele gevallen (met hoge stromingssnelheid) wordt de symmetrie gebroken, daaruit leiden we af dat voor een correcte voorspelling van de (in-) homogeniteit van het depositie proces een volledig 3 dimensionale simulatie noodzakelijk is. De virtuele reactor stelt ons ook in staat om ook de film *degradatie* processen te bestuderen alsmede de trends in systeemrespons gegeven variaties van de fysische en

chemische parameters. Zo konden we de mechanismes duiden welke verantwoordelijk zijn voor de gelaagde substraat compositie tijdens de laatste fasen van de film productie cyclus.

Interessant was de observatie van een discrepantie tussen de experimenteel gevonden filmdikte en de gesimuleerde dikte voorspellingen. Additionele numerieke simulaties lieten zien dat de correcte experimentele vorm verkregen kan worden door de elektron dichtheidsverdeling, en niet de gasstroming parameters, te variëren. Hieruit concludeerden we dat een 2 dimensionaal ontladingsmodel essentieel is. Hoofdstuk 4 beschrijft de ontwikkeling en implementatie van een dergelijk model. De simulatie resultaten met dit model laten een kwalitatief goede overeenkomst zien met de experimenteel verkregen data van de film distributie en de geobserveerde dikte variatie. Bovendien verkregen we ook een goede correlatie tussen de gemeten film eigenschappen en de gesimuleerde ion en radicaal flux in de buurt van het substraat. Verder bestudeerden we de invloed van druk, temperatuur en ontladingsparameters (intensiteit en frequentie) op het PECVD proces voor een grote range van parameter instellingen. Ook keken we naar kleine veranderingen in de inlaat en uitlaat posities en namen relatief grote effecten waar op de uniformiteit en samenstelling van de groeiende film. Hieruit concluderen we dat de depositie eigenschappen werkelijk gevoelig zijn voor de *gehele* experimentele parameter ruimte. De resultaten laten zien dat de Virtuele Reactor in staat is om trends te voorspellen in de meest relevante groei parameters en gebruikt kan worden voor de bestudering en optimalisatie van de bijzonder complexe PECVD technologie.

Na de ontwikkeling en validatie van numerieke modellen voor het PECVD proces zijn we in detail gaan kijken naar de computationele performance en de gebruikersvriendelijkheid van de Virtuele Reactor (Hoofdstukken 5 en 6). De ontwerpkeuzes waren hier dat de simulaties efficiënt dienen te lopen in zowel de *capacity* als de *capability* varianten van de Virtuele Reactor. Hiertoe bestudeerden we in detail de computationele complexiteit van de PECVD simulaties en ontwikkelden nieuwe parallellisatie methoden om de numerieke processen te versnellen. In deze optimalisatie hebben we rekening gehouden met executie van de Virtuele Reactor op sterk gekoppelde (parallelle clusters) en zwak gekoppelde (grid-gebaseerde) computersystemen. Een van de grootste uitdagingen hier was het migreren van parallelle modules van homogene clusters naar heterogene grid-resources, onder behoud van parallelle efficiëntie. Om het intrinsieke load-balancing probleem op te lossen hebben we een theoretisch model ontwikkeld en een generieke loadbalancing techniek welke dynamisch de specifieke grid-resource parameters en de applicatie parameters meeneemt in het executie optimalisatie proces. Het ontwikkelde algoritme is gevalideerd op ondermeer het Russisch-Nederlandse grid (RDG) testbed.

De resource management strategie waarin de gedistribueerde componenten afgebeeld worden op de onderliggende architectuur werd geoptimaliseerd door een serie van benchmarks op verschillende resources van het RDG testbed in detail te bestuderen. De resultaten laten zien dat de algoritmes welke ontwikkeld zijn efficiënt een dynamische adaptatie van de workload op dynamisch veranderende resources ondersteunen.

Ten slotte hebben we alle numerieke en computationele methoden geïntegreerd in één enkel ‘Problem Solving Environment’ (PSE). De Virtuele Reactor werd geïmplementeerd op het grid onder gebruikmaking van de Cross-Grid infrastructuur en werd getest met een groot aantal verschillende testbed configuraties. Hierbij verkregen we

in één systeem: veilige grid-toegang, resource discovery en resource registratie, grid-data transfer, applicatie initialisatie, editing van fysische en chemische databases, parameter specificatie, job submissie, gedistribueerde simulatie en geavanceerde visualisatie. Uitgebreide experimenten met deze Virtuele Reactor PSE leidden ons tot de conclusie dat de 'core' parallelle solver efficiënt draait op een cluster met high-speed interconnectiviteit, terwijl de andere componenten van de Virtuele Reactor efficiënt gedistribueerd kunnen worden over de (zwak gekoppelde) grid resources. Hiermee maximaliseren we de totale efficiënt van de gehele samengestelde applicatie. Loadbalancing kan verkregen worden door middel van de nieuw ontwikkelde algoritmiek zodat de simulaties automatisch adaptief geoptimaliseerd worden naar resource behoefte en resource aanwezigheid.

Na een aanloop met relatief eenvoudige modellen, via nieuwe numerieke en computationele methoden heeft dit werk geresulteerd tot een geavanceerde simulatieomgeving welke intensief gebruikt wordt in een aantal internationale projecten. Recent nog heeft dit onderzoek aanleiding gegeven tot het opstarten van een drietal nieuwe internationale samenwerkingsverbanden waarin de methoden, algoritmen en computer systemen verder ontwikkeld worden. Veel werk is gedaan en veel werk is gepland en, zoals de geschiedenis van de wetenschap ons leert, het oplossen van het ene vraagstuk opent dikwijls de deur naar vele nieuwe uitdagende en spannende vraagstukken.

Chapter 9. Acknowledgements

First and foremost, I want to express my deepest gratitude to Peter Sloot –a great supervisor and a very deserving man– for giving me the opportunity to extend my horizons, for introducing me to the scientific life at large, and for sharing his amazing enthusiasm. I also thank Yuriy Gorbachev –my Russian supervisor– for sharing his deep understanding of physics, for his patience and kind character. It has been (and hopefully will be) a pleasure to work with the two of them –not only the highly esteemed professors, but most importantly very bright personalities!

Next, I would like to express my gratitude to the Promotion Committee: Sergey Alexandrov, Marian Bubak, Wim Goedheer, Alfons Hoekstra, Jean-Pierre Taran, and Robert Meijer, for fruitful discussions and helpful comments on this manuscript. I had the honour to collaborate with you in various projects, and I am happy that you will see some of the results of this joint work included in this thesis.

Special thanks go to Jean-Pierre Taran for his inspiration, energy and remarkable trust in people. I sincerely appreciate his help with the manuscript. It started from the words "It's OK if your work is not perfect. If it were then you would be a Nobel Prize nominee and not a PhD candidate". At that time I believed that only a thousand times better work deserved a PhD degree... and I still do! ☺ Jean-Pierre, I also thank you and your family for the incredible hospitality!

I would like to thank Wim Goedheer with his group from FOM Institute for Plasma Physics and Yatin Rath with his lab from Utrecht University for long and fruitful collaboration. Your plasma model and experimental results were the corner-stones of my work.

I should specially mention Alfons Hoekstra –a brilliant organizer and scientist– who has been helping me in organizing the workshop on Simulation of Multiphysics Multiscale Systems during the last two years. I also thank Bastien Chopard and Yuriy Gorbachev who helped me with that for three years before Alfons.

Vladimir Korkhov deserves a big thank for his work on Grid implementation and performance testing. Together we developed the adaptive load balancing algorithm for parallel applications on the Grid; and Volodya was doing all the hard work on actual coding. A considerable part of Chapter 5 we wrote and published together. Later we introduced a hybrid resource management technique combining application-level automated load balancing with the system-level scheduler. This work (not included in the thesis) was done in close collaboration with Jacub Moscicki from CERN.

I am grateful to Mikhail Zatevakhin, Olga Maslennikova and Alexey Ignatiev for their work on the computer environment REAF from which the Virtual Reactor was later born; to Vitaly Schweigert and Irina Schweigert for their help with the first plasma simulations; to Dmitry Malashonok, Sergey Romanov, Maxim Kluev, Irina Shoshmina and Alfredo Tirado-Ramos for supporting the Grid testbed and assisting in our work on Grid implementation. Hans Ragas and Michael Scarpa are gladly acknowledged for their help with the 3D virtual reality visualization. The help of the emerging middleware for Grid

computing developed by the CrossGrid consortium has proven to be crucial for the success of Grid implementation.

I am lucky to work in the Section Computation Science with the very nice people and strong experts. I would like to thank Dick van Albada for his amazing attention to detail, for his help in the workshop organization, and for sharing his office during some of my visits. I did not work together with all the group members, but I always find communication with all of you very inspiring: Rob Belleman, Breanndán Ó Nualláin, Jaap Kaandorp, Simon Portegies Zwart, Elena Zudilova-Seinstra, Drona Kandhai, Bas van Vlijmen, Kasia Rycerz, Alfredo Tirado-Ramos, Alessia Gualandris, Michael Scarpa, Lilit Axner, Maxim Yurkin, Syed Murtaza, Lev Naumov, Dmitry Vasyunin, Ilkay Altintas, Derek Groen. I should also acknowledge Mike Lankamp for his superb technical support of the conference web server. I apologize if I missed some names; that must be only because I do not see you all too often.

I am indebted to Erik Hitipeuw for his enormous help with all kind of official matters and arrangements. If this thesis exists in print, it is thanks to Erik who was my proxy at the publisher. The complete list of things that Erik helped me with would be too long, so I leave this paragraph unfinished, but solely devoted to you –an indispensable irreplaceable incredible secretary, the key person in maintaining order and nice spirit of the group.

My work at the former Institute for High Performance Computing and Data Bases of St. Petersburg State Polytechnic University gave me an valuable research and life experience. I would like to thank the people I worked with: Mikhail Zatevakhin, Alexey Ignatiev, Olga Maslennikova, Elena Stankova, Maria Pavlova, Evgeniy Sokolov, Alexander Samsonov, Maria Samsonova, Alexander Boukhanovsky, Alexey Porubov, Ashot Gevorkyan, Armen Grigoryan, Denis Moskvin, Alexander Bogdanov, and many others.

I shall thank all my teachers, from my primary school to the university, for sharing their knowledge and inspiration. I'd like to specially mention my teachers: Viktor Rodionov in physics, Igor Sohan in math, and Vladimir Volovik in chemistry (St. Petersburg Phys-Math Lyceum 239). I am very grateful to Mikhail Streletz and Mikhail Shur for giving me the first research experience at the State Institute for Applied Chemistry, where I worked towards my Master thesis. Mikhail Streletz then gave me one advice that brought me here: to study English. I am happy and grateful I did follow it, and I consequently thank also my English teachers: Gillian Zakharova from Liverpool and Albina Semenovna from my primary school.

Here is a line to my dear friends from the happy high-school times: Anna Rabkina, Yuriy Kats, Denis Alekseev, Viktor Muraviov, and Andrey Ignatov. We do not meet often lately, but I carry you in my heart and always remember what Anna (a Doctor of Medicine practicing in the States) once said: "You must get a PhD. I want to have a PhD friend!" An'ka, you were the only one who ever phrased it like this, so I did it partly for you. ☺

I thank my parents and grand parents for raising me with the idea that intellectual strength is one of the most important virtues, right after the honor. I thank my late father for solving cross-words and puzzles with me, as early as I started to talk (or maybe earlier, I do not recall well). I thank my mom for her easy loving personality, for her sweet care and for being proud of me. My fascination with math and computers started in the kindergarten

(when computers did not actually exist): my grandmother was occasionally taking me to her work in a bank, and I happily struck complicated balances with astronomical numbers. At the age of 5 or 6 I was declared to have sufficient qualification to work in the bank, so I decided that calculating things could be my future job –along with the job of a detective, traveler and ballerina. (You must be laughing at this point, but my mom was an amateur ballet dancer, so I sometimes walked that special way, hoping that someone might suspect I am a ballerina. Now I hope nobody did! ☺). A programmable calculator I got as a present a few years later made me forget about the ballet and detectives: simulating the motion of stars and sledges and forests and Bob knows what was so exciting that whatever I did in the future was inclined to computer simulations. (Now I am amazed that it was possible without a 4 GHz processor and a 500 Gb hard drive!)

My last and my biggest thanks go to my family: my husband Alexander, always so patient and kind, and to my lovely daughter Sveta (translated from Russian as "light"), who grew up so smart and mature that I started taking lessons from her. I cannot thank you enough for your support and comfort, and just for being with me. Together with my mom, you are the light of my life!

Chapter 10. Publications*

1. V.V. Korkhov, V.V. Krzhizhanovskaya, J.T. Moscicki. Dynamic Workload Balancing of Parallel Applications with User-Level Scheduling on the Grid. Accepted for publication in the Future Generation Computer Systems. 2008
2. V.V. Korkhov, J.T. Moscicki and V.V. Krzhizhanovskaya. User-Level Scheduling of Divisible Load Parallel Applications with Resource Selection and Adaptive Workload Balancing on the Grid. Accepted for publication in the IEEE Systems Journal, Special Issue on Grid Resource Management. 2008
3. J.K. Rath, A. Verkerk, M. Brinza, R.E.I. Schropp, W.J. Goedheer, V.V. Krzhizhanovskaya, Y.E. Gorbachev, K.E. Orlov, E.M. Khilkevitch, A.S. Smirnov. Gas phase considerations for the deposition of thin film silicon cells by VHF-PECVD at low substrate temperatures. Accepted at the 33rd IEEE PVSEC.
4. A. Verkerk, J.K. Rath, M. Brinza, R.E.I. Schropp, W.J. Goedheer, V.V. Krzhizhanovskaya, Y.E. Gorbachev, K.E. Orlov, E.M.Khilkevitch, A.S. Smirnov. Compensation of decreased ion energy by increased hydrogen dilution in plasma deposition of thin film silicon solar cells at low substrate temperatures. Accepted at the Symposium K of the EMRS spring meeting.
5. V.V. Korkhov; V.V. Krzhizhanovskaya and P.M.A. Sloot: A Grid Based Virtual Reactor: A case study of parallel performance and adaptive load balancing. Journal of Parallel and Distributed Computing, V. 68/5, May 2008, pp 596-608. http://dx.doi.org/10.1016/j.jpdc.2007.08.010
6. V.V. Krzhizhanovskaya, V.V. Korkhov, M.A. Zatevakhin, Yu.E. Gorbachev. Parallel Distributed Computing in Modeling of the Nanomaterials Production Technologies. Parallel Computing Technologies (PAVT'2008): Proceedings of international scientific conference (St. Petersburg, 28 Jan – 1 Feb 2008). Publ.: YUSU, Chelyabinsk. 2008 ISBN 978-5-696-03720-2, pp. 585-590
7. Y.E. Gorbachev, V.V. Krzhizhanovskaya, M.V. Bogdanov, A.I. Zhmakin, A.V. Kulik, M.S. Ramm. Virtual Laboratories: Problem Solving Environments for Science and Education. Proceedings of the XIV All-Russian conference Telematika'2007. June 18-21, 2007, St. Petersburg, Russia. "University Telecommunications", St. Petersburg, pp. 108-109 (in Russian) http://tm.ifmo.ru/tm2007/db/doc/get_thes.php?id=290
8. V.V. Krzhizhanovskaya (Editor). Special Issue on Simulation of Multiphysics Multiscale Systems. International Journal for Multiscale Computational Engineering. V. 5, Issue 1, 2007. Guest editorship and Preface. DOI: 10.1615/IntJMultCompEng.v5.i1. ISSN 1543-1649
9. V.V. Krzhizhanovskaya, Y.E. Gorbachev, M.A. Zatevakhin and P.M.A. Sloot, "Virtual Reactor" – Computational Environment for Multidimensional Simulation of Plasma

* Listed in reverse chronological order

Chemical Processing, Proceedings of the XV International Conference on Computational Mechanics and Modern Applied Software Systems (CMMASS-2007), 25-31 May 2007, Alushta, Ukraine, Pbl.: MAI

10. V.V. Krzhizhanovskaya and S. Sun. Simulation of Multiphysics Multiscale Systems: Introduction to the ICCS-2007 Workshop. Proceedings of the 7th International Conference on Computational Science (ICCS-2007), Beijing, China, May 27-30, 2007, Part I, in Lecture Notes in Computer Science, Vol. 4487, pp. 755-761. Springer Berlin / Heidelberg 2007. ISSN 0302-9743 (Print), 1611-3349 (Online), ISBN 978-3-540-72583-1. DOI: http://dx.doi.org/10.1007/978-3-540-72584-8_100
11. V.V. Krzhizhanovskaya and V.V. Korkhov. Dynamic Load Balancing of Black-Box Applications with a Resource Selection Mechanism on Heterogeneous Resources of the Grid. Proceedings of the Ninth International Conference on Parallel Computing Technologies (PaCT-2007), Pereslavl-Zalessky, Russia, September 3-7, 2007, in Lecture Notes in Computer Science, Vol. 4671, pp. 245–260, Springer-Verlag Berlin Heidelberg 2007. 0302-9743 (Print) 1611-3349 (Online). http://www.springerlink.com/content/b070j64r30916320/ DOI: 10.1007/978-3-540-73940-1
12. V.V. Krzhizhanovskaya. Software Environment for Simulation of Plasma-Chemical Deposition Reactors on Heterogeneous Computational Resources of the Grid. Proceedings of the conference on Computer Technologies in Modern Researches, Polytechnic Symposium "Young Scientists to the industry of Northwest Russia". December 2006. Publ: St. Petersburg State Polytechnic University 2006, pp. 81-82. ISBN 5-7422-1365-4. Best presentation award.
13. V.V. Krzhizhanovskaya, V.V. Korkhov, A. Tirado-Ramos, D.J. Groen, I.V. Shoshmina, I.A. Valuev, I.V. Morozov, N.V. Malyshkin, Y.E. Gorbachev, P.M.A. Sloot. Computational Engineering on the Grid: Crafting a Distributed Virtual Reactor. Second IEEE International Conference on e-Science and Grid Computing (e-Science'06), 2006, pp.101. IEEE CS Press. http://doi.ieeecomputersociety.org/10.1109/E-SCIENCE.2006.42 (full paper on CD)
14. М.В. Богданов, Ю.Е. Горбачев, В.В. Кржижановская, А.В. Кулик, М.С. Рамм, Д.Х. Офенгейм, Д.Е. Тараканов, А.В. Цырюльников. " Процессы тепло- и массообмена в технологии современных полупроводниковых материалов". Методическое пособие по выполнению вычислительных лабораторных работ. С.-Петербург. 2006 (68 с.)
15. V.V. Krzhizhanovskaya, B. Chopard, Y.E. Gorbachev (Eds.) Simulation of Multiphysics Multiscale Systems. Special Issue of International Journal for Multiscale Computational Engineering. V. 4, Issue 3, 2006. Guest editorship. DOI: 10.1615/IntJMultCompEng.v4.i3, Preface: pp. 305-306 DOI: 10.1615/IntJMultCompEng.v4.i3.10. ISSN 1543-1649
16. V.V. Korkhov, V.V. Krzhizhanovskaya. Workload Balancing in Heterogeneous Grid Environment: A Virtual Reactor Case Study. Proceedings of the Second International Conference "Distributed Computing and Grid Technologies in Science and Education". Publ: JINR, Dubna, 2006, D11-2006-167, pp. 103-113, ISBN 5-9530-0138-X

17. V.V. Krzhizhanovskaya and V.V. Korkhov. Problem-Solving Environments for Simulation and Optimization on Heterogeneous Distributed Computational Resources of the Grid. Proceedings of the Third International Conference on Parallel Computations and Control Problems PACO '2006, Moscow, Russia, October 2-4, 2006. Publ: Moscow, V.A. Trapeznikov Institute of Control Sciences RAS, 2006. pp. 917-932. ISBN 5-201-14990-1.
18. V.V. Krzhizhanovskaya, V.V. Korkhov, P.M.A. Sloot. Virtual Reactor: a distributed computing environment for simulation of plasma chemical processes on heterogeneous resources of the Grid. All-Russian conference "Scientific services on the Internet: Parallel Programming Technologies". Novorossiysk, Russia, 18-23 September 2006
19. Y.E. Gorbachev, A.I. Zhmakin, M.A. Zatevakhin, V.V. Krzhizhanovskaya, M.V. Bogdanov, A.V. Kulik, D.H. Ofengeim, M.S. Ramm. From Electronic Textbooks to Virtual Laboratories. Telecommunications and Informatization in Education, N 5 (36), 2006. Publ: SGU, Moscow (in Russian)
20. V.V. Korkhov, V.V. Krzhizhanovskaya. Workload Balancing in Heterogeneous Grid Environment: A Virtual Reactor Case Study. Book of abstracts of the Second International Conference on Distributed Computing and Grid Technologies in Science and Education. Publ: JINR, Dubna, 2006, p. 93. ISBN 5-9530-0117-7. http://grid2006.jinr.ru/programme29.asp
21. V.V. Korkhov and V.V. Krzhizhanovskaya. Benchmarking and Adaptive Load Balancing of the Virtual Reactor Application on the Russian-Dutch Grid. Proceedings of the 6th International Conference on Computational Science (ICCS 2006), Reading, UK, May 28-31, 2006, Part I, in Lecture Notes in Computer Science, Vol. 3991, pp. 530 - 538. Springer Berlin / Heidelberg 2006. ISBN: 3-540-34379-2, DOI: 10.1007/11758501 http://dx.doi.org/10.1007/11758501_72
22. V.V. Krzhizhanovskaya, B. Chopard, Y.E. Gorbachev (Eds.) Simulation of Multiphysics Multiscale Systems. Special Issue of International Journal for Multiscale Computational Engineering. V. 4, Issue 2, 2006. Guest editorship. DOI: 10.1615/IntJMultCompEng.v4.i2. Preface: pp. 207 DOI: 10.1615/IntJMultCompEng.v4.i2.10 http://dx.doi.org/10.1615/IntJMultCompEng.v4.i2
23. V.V. Krzhizhanovskaya, P.M.A. Sloot and Y.E. Gorbachev. Modeling of Plasma Chemical Deposition and Degradation of Silicon Thin Films. In "Fracture of Nano and Engineering Materials and Structures", proceedings of the 16th European Conference of Fracture, Alexandroupolis, Greece, July 3-7, 2006 Editor E.E. Gdoutos, Springer ISBN-10 1-4020-4971-4, pp. 1347-1348
24. A.V. Porubov, G.A. Maugin, V.V. Gursky and V.V. Krzhizhanovskaya. On some localized waves described by the extended KdV equation. Comptes Rendus Mecanique, V. 333, Issue 7, July 2005, pp. 528-533 http://dx.doi.org/10.1016/j.crme.2005.06.003
25. V.V. Krzhizhanovskaya, P.M.A. Sloot and Y.E. Gorbachev. Grid-based Simulation of Industrial Thin-Film Production. Simulation: Transactions of the Society for Modeling and Simulation International, Special Issue on Applications of Parallel and Distributed

Simulation in Industry, January 2005, V. 81, No. 1, pp. 77 - 85. ISSN:0037-5497 DOI: 10.1177/0037549705051972 http://sim.sagepub.com/cgi/reprint/81/1/77

26. V.V. Krzhizhanovskaya, Y.E. Gorbachev and P.M.A. Sloot. A Grid-Based Problem-Solving Environment for Simulation of Plasma Enhanced Chemical Vapor Deposition. Proceedings of the International Conference "Distributed Computing and Grid Technologies in Science and Education". Publ: JINR, Dubna, 2004, pp. 262-271
27. V.V. Krzhizhanovskaya, Y.E. Gorbachev and P.M.A. Sloot. A Grid-Based Problem-Solving Environment for Simulation of Plasma Enhanced Chemical Vapor Deposition. Book of abstracts of the International Conference "Distributed Computing and Grid Technologies in Science and Education". 29 June – 2 July 2004, Dubna, Russia, Publ: JINR, Dubna, 2004, pp. 89-90. ISBN 5-9530-0052-9
28. Y.E. Gorbachev, M.A. Zatevakhin, Ignatiev A.A., Krzhizhanovskaya. Virtual Laboratory for Research and Education. XI International Conference on High Technologies and Quality of Education and Science, St. Petersburg Polytechnic University. 27-28 February 2004. Publ.: SPbSPU, St. Petersburg, 2004, pp. 59-60
29. V.V. Krzhizhanovskaya, M.A. Zatevakhin, A.A. Ignatiev, Y.E. Gorbachev, W.J. Goedheer and P.M.A. Sloot. A 3D Virtual Reactor for Simulation of Silicon-Based Film Production. Proceedings of the 5th International Bi-Annual Symposium on Computational Technologies for Fluid/Thermal/Structural/Chemical Systems with Industrial Applications, ASME/JSME PVP Conference. San Diego, California USA, July 25-29, 2004. ASME PVP-Vol. 491-2, pp. 59-68, PVP2004-3120. ISBN 0-7918-4686-5
30. Y.E. Gorbachev, V.V. Krzhizhanovskaya, P.M.A. Sloot. Using GRID technologies for simulation of complex physical-chemical systems. Proceedings of the XI All-Russian conference Telematika'2004. June 7-10, 2004, St. Petersburg, Russia. Publ.: SPbSU ITMO "Informika", Moscow, v. 1, pp. 165-167 (in Russian) http://tm.ifmo.ru/tm2004/db/doc/get_thes.php?id=392
31. Y.E. Gorbachev, M.A. Zatevakhin, Ignatiev A.A., V.V. Krzhizhanovskaya. Simulation of heat and mass transfer processes in PECVD reactors for silicon films growth under the conditions of intense formation of higher silanes in the gas phase. The 5th Minsk International Heat and Mass Transfer Forum, MIF-2004. May 24-28, 2004. Minsk, Belarus. Book of abstracts, V. 1, pp. 301-302 (in Russian). Proceedings published on CD, index S04/4-08. Proceedings on the Web: http://relay.itmo.by/forum/mif5/S04/4-08.PDF (in Russian)
32. Y.E. Gorbachev, M.A. Zatevakhin, V.V. Krzhizhanovskaya, Ignatiev A.A. A Software-hardware System for Simulation of Technological Devices on Distributed Computer Clusters. Proceedings of the X All-Russian conference Telematika-2003. April 14-17, 2003, St. Petersburg, Russia. Publ.: SPbSU ITMO "Informika", Moscow, v. 1, pp. 165-167 (in Russian) http://tm.ifmo.ru/tm2003/db/doc/get_thes.php?id=353
33. A.V. Porubov, G.A. Maugin, V.V. Gursky, V.V. Krzhizhanovskaya, D.V. Shevchenko. Localized nonlinear strain waves in complex dispersive-dissipative solids. Proceedings of the XXXI International Conference "Advanced Problems in Mechanics" Russia, St.

Petersburg, June 22-July 2, 2003, Publ.: RAS IPM GAMM, St. Petersburg, Russia, pp. 270- 279

34. A.V. Porubov, G.A. Maugin, V.V. Krzhizhanovskaya, and V.V. Gursky. Large Solitary Waves Formation in Microstructured Solids. Proceedings of the Euromech 436: Nonlinear Waves in Microstructured Solids. Tallinn, Estonia, May 29 – June 1, 2002, p. 26
35. V.V. Krzhizhanovskaya, M.A. Zatevakhin, A.A. Ignatiev, Y.E. Gorbachev, P.M.A. Sloot. Distributed Simulation of Silicon-Based Film Growth. Proceedings of the Fourth International Conference on Parallel Processing and Applied Mathematics (PPAM 2001) Naleczow, Poland, September 9-12, 2001, in series Lecture Notes in Computer Science, Vol. 2328, pp. 879-888. Springer-Verlag 2002. ISBN: 3-540-43792-4 , ISSN: 0302-9743 http://www.springerlink.com/content/erj5rtybb3030pe6/
36. Y.E. Gorbachev, M.A. Zatevakhin, V.V. Krzhizhanovskaya, A.A. Ignatiev, V.K. Protopopov, N.V. Sokolova, A.B. Witenberg. Distributed Simulation of Amorphous Hydrogenated Silicon Films: Numerical Experiments on a Linux Based Computing Environment. Proceedings of the International Conference on Computational Science (ICCS 2001), San Francisco, USA, May 2001, in series Lecture Notes in Computer Science, Vol. 2073, pp. 483-491. Springer-Verlag 2001. ISBN: 3-540-42232-3 ISSN: 0302-9743
37. M.A. Zatevakhin, A.A. Ignatiev, V.V. Krzhizhanovskaya, V.K. Protopopov, A.B. Witenberg, Y.E. Gorbachev, V.A. Schweigert. Simulation of Heat and Mass Transfer Processes in PECVD Reactors for a-Si:H Films Growth. Proceedings of the XI International Conference on Computational Mechanics and Modern Applied Software Systems. 2-6 July 2001, Moscow, Russia, Pbl.: MAI, pp. 191-193
38. M.A. Zatevakhin, V.V. Krzhizhanovskaya. Deep Convection: Its Role in Transporting Pollutants from the PBL to the Upper Troposphere. Proceedings of the Conference “Air transport: Impact on ozone, climate and airport air quality”. May 16-17 2001, ONERA, France, pp. 57-60
39. Y.E. Gorbachev, M.A. Zatevakhin, V.V. Krzhizhanovskaya, V.A. Shveigert. Special Features of the Growth of Hydrogenated Amorphous Silicon in PECVD Reactors. Technical Physics, Vol. 45, N 8, pp. 1032-1041. Publ.: MAIK Nauka/Interperiodika 2000. ISSN: 1063-7842
40. Y.E. Gorbachev, M.A. Zatevakhin, V.V. Krzhizhanovskaya. Influence of Silane Dilution with Hydrogen and Inert Gases on the Dynamics of Amorphous Hydrogenated Silicon Films Growth in PECVD Reactors. Proceedings of the 2nd International Conference on Amorphous and Microcrystalline Semiconductors, St. Petersburg, Russia, 2000
41. M.A. Zatevakhin, V.V. Krzhizhanovskaya, Y.E. Gorbachev. Simulation of a-Si:H Film Growth During Remote PECVD. XIV Europhysics conference on Atomic and Molecular Physics of Ionized Gases (ESCAMPIG 98), Malahide Dublin, Ireland,1998. Europhysics Conference Abstracts 1998, V. 22H, pp. 488-489. Publisher: The European Physical Society, Switzerland. ISSN 0378-2271 http://direct.bl.uk/bld/PlaceOrder.do?UIN=054438709&ETOC=RN

42. Y.E. Gorbachev, M.A. Zatevakhin, V.V. Krzhizhanovskaya. Hydrogenated Amorphous Silicon Film Growth in Remote PECVD. Proceedings of the All-Russian Symposium on Amorphous and Microcrystalline Semiconductors, St. Petersburg, Russia, 5-9 July 1998, p.24

43. Y.E. Gorbachev, M.A. Zatevakhin, V.V. Krzhizhanovskaya. Simulation of Amorphous Hydrogenated Silicon Plasma Enhanced Chemical Vapor Deposition in a Triode Reactor. Proceedings of the NATO workshop on Plasma Physics, St. Petersburg, Russia, 1997