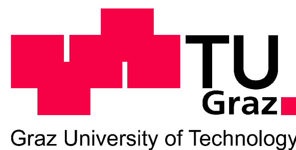
Matthias Haselmann, BSc

Optical Diagnostics of High Density Plasma Chemical Vapor Deposition Processes in Semiconductor Wafer Production

MASTER THESIS

For obtaining the academic degree
Diplom-Ingenieur

Master Programme of
Technical Physics



Graz University of Technology

Supervisor:

Ao.Univ.-Prof. Dipl.-Ing. Dr.techn. Theo Neger

Institute of Experimental Physics

Graz, November 2013

Acknowledgment

I would like to express my deep gratitude to my master thesis supervisor Prof. Theo Neger for his continuous support, patience, motivation and engagement through the learning process of this work. Furthermore I would like to thank ams AG for the opportunity to write a master thesis about this topic. I would like to thank all people who directly or indirectly helped me there to create this master thesis. Special thanks are given to my two advisors Peter Kailbauer and Günter Hayderer for their support and engagement. Moreover, I would like to thank Peter Scheibelhofer and Rolf Konrader for their input and valuable discussions. Last but not least I would like to thank my parents, who have supported me throughout my entire course of study.

Deutsche Fassung:
Beschluss der Curricula-Kommission für Bachelor-, Master- und Diplomstudien vom 10.11.2008
Genehmigung des Senates am 1.12.2008

EIDESSTATTLICHE ERKLÄRUNG

Ich erkläre an Eides statt, dass ich die vorliegende Arbeit selbstständig verfasst, andere als die angegebenen Quellen/Hilfsmittel nicht benutzt, und die den benutzten Quellen wörtlich und inhaltlich entnommenen Stellen als solche kenntlich gemacht habe.

Graz, am …………………………… ………………………………………………..
(Unterschrift)

Englische Fassung:

STATUTORY DECLARATION

I declare that I have authored this thesis independently, that I have not used other than the declared sources / resources, and that I have explicitly marked all material which has been quoted either literally or by content from the used sources.

…………………………… ………………………………………………..
date (signature)

Abstract

This work is devoted to the application of a process monitoring method of High Density Plasma Chemical Vapor Deposition (HDP CVD) of conform dense SiO_2-layers on 3d-wafer structures by using optical emission spectroscopy (OES). From the obtainable optical spectra of the relevant gaseous species contained in the reactor chamber a reliable prediction tool for the deposition process is developed. Early recognition of faults in manufacturing can save high costs in semiconductor manufacturing and prevent malfunctions of the chips. By the applied kind of plasma diagnostics, relative variations of the concentrations of the relevant gas species that are mainly produced by electron impact with the precursors SiH_4 and O_2, are observed via relative line intensity measurements of suitable spectral lines.

The information on gas species concentrations acquired this way is used in a statistical model for the prediction of the deposition rates. The comparison of a linear and non-linear model based on Gaussian Process Regression (GPR) shows that a linear model is sufficient for considering problem. Since deposition takes place on the chamber walls, too, resulting in slightly increasing deposition rates, cleaning is periodically required. Obviously the cleaning status of the process chamber has impact on the quality of the deposited layers on the wafers. So efforts are made to include this matter in the prediction model in view of a successful fault detection and classification (FDC) and virtual metrology (VM) of the whole operation of the HDP CVD machine.

This master thesis includes also an appendix giving detailed information on the collisional radiative model as well as the GPR statistical model applied within the frame of this work. The experimental work has been performed at the semiconductor manufacturer ams AG in Unterpremstätten/Austria.

Contents

1 Introduction **11**

2 Fundamentals of CVD **13**
2.1 Definition . . . 13
2.2 Classification . . . 13
2.3 Construction and process principles of a CVD system . . . 15
2.3.1 Precursor supply system . . . 15
2.3.2 Reactor chamber . . . 15
2.3.3 Effluent gas system . . . 16
2.4 Continuously stirred reactor (CSTR) approximation . . . 16
2.5 Mass transport phenomena in gaseous media . . . 17
2.5.1 Diffusive mass transport . . . 18
2.5.2 Convective mass transport . . . 21
2.5.3 Plug and Poisseuille flow . . . 21
2.5.4 Transport regimes . . . 22
2.6 Reaction chemistry . . . 25
2.6.1 Thermodynamic approach . . . 25
2.6.2 Kinetics . . . 26
2.6.3 Surface reaction kinetics . . . 30
2.6.4 Film conformality . . . 32

3 Plasma enhanced CVD **35**
3.1 Ambipolar diffusion . . . 35
3.2 Capacitively driven plasmas . . . 37
3.2.1 Introduction . . . 37
3.2.2 Construction and principle . . . 37
3.2.3 Impact of the excitation frequency to the plasma properties . . . 39
3.2.4 Sheath model . . . 42
3.2.5 Sheath ion flux . . . 44
3.2.6 Plasma ignition: Paschen's law . . . 46
3.3 Inductively driven plasmas . . . 46
3.3.1 Construction and principle . . . 46
3.3.2 Magnetic skin effect . . . 48
3.3.3 Uniformity . . . 48
3.3.4 Plasma ignition . . . 49
3.4 Microwave driven plasmas . . . 49
3.4.1 Excitation mechanism and efficiency . . . 50
3.5 Plasma damage . . . 50
3.5.1 Plasma charging due to a lateral non-uniformity in the DC plasma potential . . . 50
3.5.2 Damage from impinging particles and photons . . . 50
3.5.3 Arcing . . . 51

4 High density plasma CVD **53**
4.1 Principle 53
4.2 Construction and operation 55
4.2.1 Combined ICP and CCP power sources 55
4.2.2 Wafer temperature and cooling 56
4.2.3 Film uniformity 56
4.2.4 Gas flow and pumps 56
4.2.5 Chamber wall and cleaning procedure 57
4.3 Transport analysis 57
4.3.1 Volumetric flow 57
4.3.2 Average gas residence time and velocity 58
4.3.3 Mass diffusivity 58
4.3.4 Surface reaction rate constant 59
4.3.5 Average consumption time 60
4.3.6 Transport limited deposition rate 60
4.3.7 Transport analysis summary 60
4.4 Process chemistry 61
4.4.1 Dominant gas phase reactions 61
4.4.2 Dominant surface reactions 62

5 Optical emission spectroscopy of HDP CVD processes **65**
5.1 Motivation 65
5.2 Experimental setup and procedure 66
5.3 Optical spectrum of the HDP CVD process 66
5.4 Optical spectrum of the cleaning procedure 72
5.5 Time course of the line-intensities 72
5.6 Determination of some relevant plasma parameters 73
5.6.1 Gas temperature 74
5.6.2 Ion density 74
5.6.3 Electron density 75
5.6.4 Electron temperature 75
5.6.5 Simultaneous determination of electron density and electron temperature by line-intensity ratios 76
5.6.6 Plasma species concentrations 76
5.7 Statistical prediction of the deposition rate 77
5.7.1 Influences of the cleaning cycles 77
5.7.2 Non-linearity of the plasma parameters 78
5.7.3 Data for modeling 78
5.7.4 Linear modeling 78
5.7.5 Non-linear modeling by GPR 81
5.7.6 Summary 81

6 Conclusion **83**

A Collisional radiative model (CRM) **87**
A.1 Introduction 87
A.2 CRM-model with consideration of the first 14 excited states of argon in an argon plasma 87

A.3 Application to a pure argon plasma 92
A.4 Application to the HDP CVD plasma 93

B Gaussian process regression (GPR) 95

1 Introduction

A permanently growing cost pressure and continuously increasing requirements for the production efficiency demand innovative and pioneering process technologies in semiconductor industry. This leads to a high level of automation of the individual manufacturing processes. A wafer has a high degree of complexity regarding the necessary process steps. About 500 or even more single processes have to be done. It is not astonishing that even small deviations of one of these processes from regular operation can cause malfunctions of the chips on the wafers. In order to verify the product's quality, nowadays control measurements are done randomly after each individual process. However, these control measurements represent a significant cost factor. In order to reduce them, systems like Fault Detection and Classification (FDC) and Virtual Metrology (VM) can be implemented. They allow a faster detection and classification of machine faults and - in the best case - a reliable prediction of process results. A reduced frequency of control measurements is generally accompanied by a reduction of the necessary cycle time, resulting in a probably significant cost advantage in view of the large numbers of wafers produced by a machine.

For appropriate FDC and/or VM suitable process data are required. However, the data of the machine itself are frequently not sufficient, so additional sensors have to be added. The High Density Plasma Chemical Vapor Deposition (HDP CVD) process is no exception in this respect. Moreover, plasma processes are rather critical and difficult to be controlled. They have a strong impact on the product's quality and can go wrong pretty easy.

The HDP CVD process is a special type and step in semiconductor manufacturing. $\mathrm{SiO_2}$ is conformly deposited on a 3d-wafer structures, typically on lines of aluminum. Conform deposition means that gaps with an high-aspect ratio, which arise automatically because of the metallization, are filled without voids.

Since usually not only a deposition in vertical direction, but also in horizontal direction takes place, this is not a trivial problem to be solved. When the gap overgrows firstly in horizontal direction, the probability is high that voids appear underneath. This may have a massive impact to the electrical properties.

In the HDP CVD process voids can be avoided by simultaneous ion sputtering with a maximum efficiency at 45%-inclined surfaces. Thus, critical edges of gaps are prevented from growing together. The second reason why HDP CVD is used for deposition on lines of aluminum is their temperature sensitivity. Since aluminum melts at about $660°\mathrm{C}$ the gas temperature of the process must be kept below. HDP CVD has a low pressure (about $5\,\mathrm{mTorr}$) and a gas temperature of about $250 - 400°\mathrm{C}$. Furthermore, it works in a plasma state with hot electrons up to $50000°\mathrm{C}$, which cause the decomposition of the precursor gasses necessary for deposition.

VM is not possible with machine data only. Therefore an Optical Emission Spectrometer (OES) was added as sensor to monitor to state of the plasma in this work. The main task of this master thesis was to investigate the potential of an OES-sensor for the HDP CVD machine with respect to FDC and VM. The advantages of it are quite convincing, apart from the fact that it is the most obvious way to monitor the state of a plasma. On the one hand it is a completely non-invasive method, it is fixed on the reactor outside of its window. On the other hand, it is comparatively cheap and easy to be implemented.

This work comprises an overview about Chemical Vapor Deposition (CVD) in general, followed by a more detailed part about Plasma Enhanced CVD (PECVD) processes, both oriented mainly on the work of Zuraw and Dobkin [1]. As a further consequence a detailed discussion of the certain HDP CVD process is presented. The final part describes results of the optical emission spectroscopy of HDP CVD processes. The focus here is the prediction of the thickness of the deposited SiO_2 layer. For that a suitable statistical model (Gaussian Process Regression) for machine learning was applied, which is presented, too. It is non-parametric and thus can be also used for mapping non-linear relationships.

This work was performed within the cooperation of the Graz University of Technology and the semiconductor manufacturer ams AG.

2 Fundamentals of CVD

2.1 Definition

Chemical Vapor Depositions (CVD) are chemical processes within an activated environment (heat, light, plasma) that are used for the production of highly pure and uniform solid materials with reliable reproducibility at reasonable high deposition rates [2]. By inflowing precursor gases into the reactor chamber, which decompose mainly into radicals within the homogeneous gas phase, a solid component is finally deposited by heterogeneous chemical reactions on/near the surface of the usually heated substrate. This happens, of course, if volatile compounds of the layer components as precursor gases are assumed, which start depositing under a certain reaction temperature. Therefore the substrate is preferably heated separately. In order to reduce undesired gas phase reactions which lead i.a. to the formation of solid particles in the gas phase, the CVD process is performed at reduced pressure. The produced by-products are pumped out together with unused parts of the precursor gases. One of the main advantages of chemical vapor deposition (CVD) is its conform deposition, which also makes possible the deposition of complex 3-dimensional structures and shapes compared to physical vapor deposition (PVD). So the range of producible materials is enormous and ranges from ultra-fine powders and diamonds up to highly pure and dense thin films on nano-structured surfaces, as they are needed in semiconductor manufacturing. [2] [3]

2.2 Classification

Since CVD is that versatile there are many classifications possible:

Classification by pressure:

- APCVD (Atmospheric pressure CVD)
- RPCVD (Reduced pressure CVD): Reduced operating pressure range $1-100\,\mathrm{hPA}$.
- LPCVD (Low-pressure CVD): Operating pressure range 0,1-1000 Pa.
- UHVCVD (Ultra-high vacuum CVD): range is typically below $10^{-6}\,\mathrm{Pa}$.

As already mentioned a reduced pressure minimizes undesired gas phase reactions which lead i.a. to the formation of solid particles in the gas phase. Furthermore it provides a higher diffusivity of reactants to the substrate and consequently fewer defects in the deposited layer, better conformity and a better film uniformity [4].

Classification by the stimulation of the chemical reactions: The first three techniques are associated to plasma processing:

- MPCVD (Microwave plasma-assisted CVD): The precursors are decomposed by strong microwave radiation, mostly before they are injected into the reactor. Therefore deposition at lower temperatures is possible.

- PECVD (Plasma enhanced CVD): Electron energy that goes along with the plasma state is responsible for the decomposition of the precursors. The plasma state is reached usually by inductively or capacitively coupled radio-frequency (RF) radiation or simply by electrical power at high voltage. This leads to gas breakdown and results in a glow discharge plasma. It requires a reduced pressure, usually in the low pressure region, which enables electrons to reach extremely high temperatures of about 20000 K to 50000 K while the actual gas temperature is much lower (500-1000 K). Therefore PECVD can be used for temperature sensitive substrates as i.e. wafers.

- RPECVD (Remote plasma enhanced CVD): Inductively or capacitively coupled radio-frequency (RF) radiation is used, but the wafer is not directly placed in the plasma region. This enables a further reduction of the operating temperature down to room temperature.

- PICVD (Photo-initiated CVD): Ultraviolet-light is used for the decomposition of the precursors and the stimulation of chemical reactions.

- CCVD (Combustion CVD): The chemical reactions of this CVD process is stimulated using flames.

- CVD processes do not need explicit stimulation of chemical reactions in the gas phase. Frequently only a heating of the substrate is provided to cause the formation of the desired layer. An example is MOCVD (Metalorganic CVD). In this case metalorganic gases are used as precursors, which are volatile at relatively low temperatures.

The classifications presented so far are probably the most important. Further classifications are:

Classification by physical properties of the precursors: Beside classical CVD in which the precursors are gases, they also can be liquid or non-volatile. In this case there exist two possibilities for injection:

- DLICVD (Direct liquid injection CVD): The precursors are directly injected in a liquid form or as appropriate solvent in which they are dissolved. After that the injected liquid is vaporized and transported to the substrate in the same way as in classical CVD.

- AACVD (Aerosol assisted CVD): The precursors are injected in form of an aerosol, which i.e. can ultrasonically be generated.

High or low operating temperature: The operating temperature is frequently critical since many substrates are temperature sensitive and may be damaged or destroyed above a certain temperature. Especially the plasma-driven variants of CVD provide lower temperatures and are therefore preferably used for semiconductor manufacturing, since i.e. aluminum conductor tracks melt above about $660°\mathrm{C}$, which is easily reached by conventional downstream CVD processes.

Hot or cold wall design: If the walls of the reactor are heated one speaks of a hot wall design. Cold wall design then corresponds to unheated walls [4].

Closed or open CVD system: In a closed system the chemicals are packed into the reactor and are not supplemented or removed until the end of the deposition process.

Additionally, a certain temperature may be applied [5] to ensure that the reaction is running into the proper direction, since reversible chemical reactions may occur. In comparison to that open CVD systems provide continuous passing in of the reactor gas and removing of reactants and by-products, respectively. Of course, removing of the by-products only is not possible. Anyway, removing of the reactor gases has to be done in order the keep a constant gas composition and pressure in the reactor chamber. Therefore, the open design is preferably used and is exclusively applied in semiconductor manufacturing.

Nevertheless, there are some CVD processes which can not be assigned exactly to one of these categories such as the HPCVD (Hybrid Physical CVD) that combines vaporization of a solid source with chemical decomposition of precursor gases.

2.3 Construction and process principles of a CVD system

Although CVD systems can be very different, depending on deposited material and substrate geometry etc., their construction can be roughly divided into three main components: [2]

- precursor supply system
- reactor chamber
- effluent gas system

2.3.1 Precursor supply system

The task of this unit is to provide the reactor chamber with the precursor. Depending on their aggregate state and their volatility the precursor supply system has to fulfill different requirements. The simplest case would be if all precursors are gaseous. The gases are then provided by high pressure gas reservoirs and are injected by nozzles and led through flow meters in order to control amount injected. In many cases the gases are purified further to reduce certain contaminants which come with the gas transfer lines.

If the precursor are in a liquid form they are often vaporized by a blubber or flash evaporation and then transported by reactive or inert gases to the reactor chamber. The injected amount can be calculated by knowing the vapor pressure and its temperature as well as the carrier gas flow rate. Solid precursors can be either sublimated or solved in an appropriate solvent and then be treated in the same way as liquid precursors.

2.3.2 Reactor chamber

The reactor chamber of the CVD system usually consists of a susceptor (substrate holder) as well as a handling system for the placement of the substrates. One of the main tasks of the reactor chamber is to bring the substrate up to the desired temperature and to hold it constant there. It is important to ensure a proper temperature uniformity, otherwise the deposition uniformity may significantly be affected.

Furthermore, the walls are heated in a hot wall reactor design mainly to prevent particles from falling down to the substrate, mainly generated by temperature induced

stress. Since the deposition process preferably heats the walls, it is therefore the better way to continue the heating by separate heaters when no process is running. Nevertheless, the chamber has to be cleaned periodically, because the particle problem generally aggravates the thicker the deposited layer on the wall gets.

Depending on the application and its precursors the chamber may also need a certain power supply (i.e. inductively/capacitively coupled RF-radiation or UV-light, etc.) for any decomposition and reactions of the precursors. Uniformity of the precursor distribution and of its decomposition is important again for the deposition uniformity. For a satisfying precursor distribution the design and arrangement of the gas injecting nozzles are responsible. A uniform decomposition of the precursors is enabled by a well-designed power supply for the gas decomposition.

2.3.3 Effluent gas system

The effluent gas system consists of a pumping unit for removing the by-products and holding a constant gas pressure in the reactor chamber to ensure consistent reaction conditions. The type of the pump depends on the pressure region of the process. In the low pressure region and above, a rotary vane pump is sufficient. In the high and ultra-high vacuum region a downstreamed turbomolecular pump is necessary.

Moreover, the effluent gas system has to ensure that the mostly toxic, hazardous and/or polluting reactor gases are removed safely. In order to prevent damage to the pumping system, corrosive gases (i.e. HF) have to be neutralized or trapped before entering it. Inflammable gases (i.e. H_2) are burned off. Very expensive ones are usually recycled and reused.

2.4 Continuously stirred reactor (CSTR) approximation

The continuously stirred reactor (CSTR) approximation [6] provides a zero-dimensional transport analysis of a reactor. The concentrations can be considered to be constant everywhere. The basic idea of the approach is depicted in figure 1. The model is only valid under the condition that the mixing process is much faster than the consumption of the reactants as well as their removal.

Figure 1: Schematic view of zero-dimensional reactor model (also known as Continuously Stirred Reactor approximation (CSTR)) (after [6]). n_{in}, n_{ch} and n_{out} are the precursor densites at the gas inlet, in the chamber and at the gas outlet, respectively. A marks the area where deposition with a surface reaction rate constant K_s takes place.

By considering the ratio between used reactants and streamed-in reactants it can be distinguished between starved or differential reactors: The relative consumption of the reactants can be expressed as

$$X = \frac{\Phi_{in} - \Phi_{out}}{\Phi_{in}} \quad (1)$$

where Φ_{in} and Φ_{out} are the inlet and the outlet flow of the reactor chamber. Basically, the amount of reactants entering the chamber has to be equal to the amount of reactants leaving the chamber plus the surface reaction rate $R = K_s A\, n_{ch}$,

$$\Phi_{in} n_{in} = \Phi_{out} n_{out} + K_s A\, n_{ch}, \tag{2}$$

where n_{in}, n_{out} and n_{ch} are the inlet, outlet and chamber precursor concentrations. Since this is a 0-dimensional model n_{out} has to be the same as n_{ch}. Furthermore it is assumed that the outlet gas flow Φ_{out} equals approximately the inlet gas flow Φ_{in}, since normally by-products arise in addition to the deposition. In combination with the average precursor consumption time

$$t_{\text{cons}} = \frac{V}{K_s A} \tag{3}$$

and the average precursor residence time

$$t_{res} = \frac{V}{\Phi_{in}}, \tag{4}$$

the chamber precursor concentration n_{ch} is

$$\Phi_{in} n_{in} = (\Phi_{in} + K_s A)\, n_{ch} \Longrightarrow n_{ch} = \frac{\Phi_{in} n_{in}}{\Phi_{in} + K_s A} = \frac{n_{in}}{1 + \frac{t_{res}}{t_{\text{cons}}}}. \tag{5}$$

If $t_{res} << t_{\text{cons}}$ and X is close to zero, respectively, the chamber precursor concentration is actually equal to the inlet precursor concentration and will also be very constant and uniform along the whole chamber. This situation requires a surface limited reaction rate and is called a *differential reactor*. On the basis of the uniform reactant concentration along the whole chamber a solid uniformity is accomplished assuming that the substrate has a uniform temperature.

If $t_{\text{cons}} << t_{res}$ or X is close to one, it is called a *starved reactor* that obviously provides large surface reaction rates. The deposition rate is mass transport limited then. However, the uniformity is much worse if no clever arrangements of injectors are found or a location-dependent excitation of precursors can be provided. The differential reactor type is often used for research purposes, whereas the starved reactor is rather used in industrial applications.

2.5 Mass transport phenomena in gaseous media

Since at every CVD process precursors have to be transported from the inlet to the substrate to preserve deposition, mass transport is a necessary property. In case of a mass transport limited CVD process the mass transport rate influences heavily the deposition rate [6]. In that case mass transport is also responsible for the uniformity, which is very important for many practical applications [2]. In general mass transport phenomena can roughly be distinguished by two contributions [2]:

- Mass transport of the reactants along the reactor chamber.
- Mass transfer of reactants close to the surface of the substrate. This includes adsorption of the reactants by the substrate surface as well as any desorption of the by-products.

Overall, two mechanisms are responsible for mass transport within the gas phase, if the interaction with surfaces is excluded. Firstly convection, which results from the bulk motion of the gas and can be described by fluid dynamics. The flow is assumde to be incompressible, because its velocity u is much smaller than the velocity of sound and therefore smaller than the velocity of the individual molecules. Secondly diffusion resulting of the thermal motion of the particles in the gas.

2.5.1 Diffusive mass transport

A net mass transport of one particle type due to diffusion takes place only if the gradient of concentration of the considered particle type is non-zero. The flux due to diffusion can be described by the first law of Fick. It states that the flux is proportional to the concentration gradient and a specific diffusion constant D,

$$J_{\text{diff}} = -D\,\frac{\partial n}{\partial x}. \tag{6}$$

Note that this equation and the following one are assuming one spatial dimension and can be easily extended to more. With the help of the continuity equation

$$\frac{\partial n}{\partial t} = -\frac{\partial J_{\text{diff}}}{\partial x} \tag{7}$$

the second law of Fick can be obtained,

$$\frac{\partial n}{\partial t} = \frac{\partial}{\partial x}\left(D\,\frac{\partial n}{\partial x}\right). \tag{8}$$

If D is constant with respect to spatial variations, which is actually the case in modern CVD reactors because of its homogeneity of the gas phase, the second law of Fick can be simplified to

$$\frac{\partial n}{\partial x} = D\,\frac{\partial^2 n}{\partial x^2}. \tag{9}$$

Assuming a fixed concentration $n(0)$ at the position $x = 0$ and an initial concentration of zero at $x > 0$ the equation above has the following solution

$$n(x,t) = n(0)\,\text{erfc}\left(\frac{x}{2\sqrt{D\,t}}\right), \tag{10}$$

(erfc error function). The term $2\sqrt{D\,t}$ is known as the diffusion length L_D and is generally characteristic for diffusion problems. As time t usually the average residence time t_{res} of gases in the reactor is used, which can be estimated by the average flow velocity U and the length L of the system that corresponds approximately the way that the precursors have to go to reach the outlet,

$$L_D = \sqrt{4D\,t_{res}} = \sqrt{\frac{4DL}{U}}. \tag{11}$$

The diffusion length L_D, which corresponds obviously to the average distance traveled by a particle, is often compared to relevant parameters of the system such as the length

of the system L [1]. Two special cases have to be distinguished, namely whether the diffusion length is much longer or much shorter than the length of the system.

Assuming a steady state in the first case ($L_D >> L$) means that the profile of concentration is kept constant or at least slightly decreasing linearly due to diffusion. Therefore fluid velocity would not be required for a uniform distribution of gases. However, in the second case ($L_D << L$) the profile of concentration can have a large gradient assuming negligible convection, because the gases are removed before they can fully spread out due to the limited residence time. Moreover, abrupt changes in concentration within one region of the reactor would have no impact to other regions that are sufficiently far away [6].

Another common way to describe the influence of diffusion in a reactor system is the Peclet number Pe, which is defined as [6]

$$Pe = 4\left(\frac{L}{L_D}\right)^2 = 4\frac{L^2U}{4DL} = \frac{LU}{D}. \tag{12}$$

The first case $L_D >> L$ according to a diffusion dominated concentration profile corresponds to $Pe << 1$ and vice versa. The Peclet number shows very clearly the three parameters that are critical for the diffusion properties of the reactor.

So far fixed boundary conditions were assumed that actually do not correspond well to the requirements of a CVD system, since deposition is not taken into account at all. The next step is to consider the case that species are consumed at the boundary. If the reactor is assumed to be long and thin, the gradient in concentration across the region can be neglected. However, there is such one along its length. Assuming steady state, the net flux into the reactor has to balance the loss due to deposition [1].

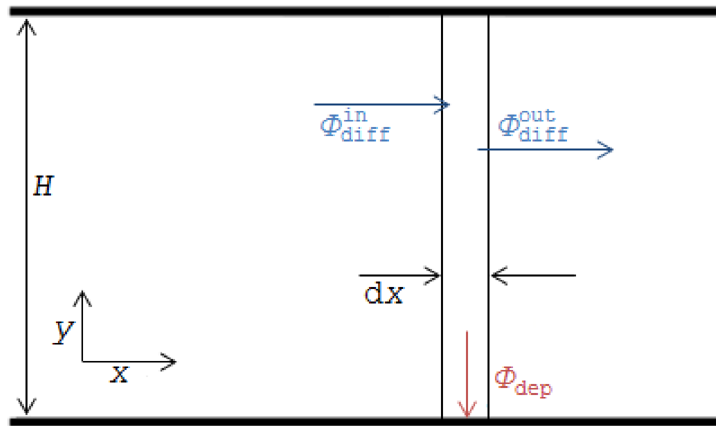


Figure 2: Schematic view of a long and thin reactor in the Thiele transport regime. The precursor concentration n is constant in the infinitesimal volume with the thickness dx.

The net flux into the region $x \rightarrow x + \mathrm{d}x$ due to diffusion is according to first law of Fick

$$\begin{aligned} J_{\mathrm{diff}} = J_{\mathrm{diff}}^{in} - J_{\mathrm{diff}}^{out} = -D\frac{\mathrm{d}n}{\mathrm{d}x}\|_x + D\frac{\mathrm{d}n}{\mathrm{d}x}\|_{x+\mathrm{d}x} = \\ = D\left[n'(x+\mathrm{d}x) - n'(x)\right] = D\left[\frac{n'(x+\mathrm{d}x)-n'(x)}{\mathrm{d}x}\mathrm{d}x\right] = D\frac{\mathrm{d}^2n}{\mathrm{d}x^2}\mathrm{d}x. \end{aligned} \tag{13}$$

Consequently, the net diffusion flux into the marked volume is

$$\Phi_{\text{diff}} = ZHJ_{\text{diff}} = ZHD\,\frac{\mathrm{d}^2 n}{\mathrm{d}x^2}\,\mathrm{d}x. \tag{14}$$

On the other hand the net flux out of the marked volume due to deposition is

$$\Phi_{\text{dep}} = Z\,\mathrm{d}x\,R = Z\,\mathrm{d}x\,K_s\,n, \tag{15}$$

where R is the deposition rate with $R = nK_s$, assuming that deposition only occurs on the bottom surface. By equalizing Φ_{diff} and Φ_{dep} one gets

$$\Phi_{\text{diff}} = \Phi_{\text{dep}} \Rightarrow ZHD\,\frac{\mathrm{d}^2 n}{\mathrm{d}x^2}\,\mathrm{d}x = Z\,\mathrm{d}x\,K_s\,n \Rightarrow \frac{\mathrm{d}^2 n}{\mathrm{d}x^2} - \frac{2K_s}{HD}\,n = 0. \tag{16}$$

The solution of this differential equation is

$$n(x) = n(0)\,e^{x/L_{th}}, \quad L_{th} = \sqrt{\frac{HD}{K_s}}, \tag{17}$$

where L_{th} is called the Thiele length and describes how fast the concentration drops along the reactor. If it is small compared to the reactor length L the concentration drops down quickly near the inlet at $x = 0$ and consequently not much deposition occurs at larger values of x [1]. However, if the Thiele length L_{th} is big compared to the reactor length L, the concentration is very uniform over x and thus the deposition rate. Therefore, the Thiele length can serve as an indicator for the longitudinal uniformity in diffusion dominated systems. A good example is a tube reactor with closely-spaced wafers. Since diffusion usually dominates in between these wafers the Thiele length has to be much larger than the wafer size in order to reach a satisfying uniformity.

By dividing the length of the system L by the Thiele length L_{th} the so called Thiele modulus is obtained,

$$\theta = \frac{L}{L_{th}} = \sqrt{\frac{K_s L^2}{HD}}. \tag{18}$$

The reactor dimensions are usually specified by the application and can therefore not be changed a lot. Therefore, the best possibility to adjust the Thiele modulus offers the diffusion constant D, which can be done i.e. by changing the chamber pressure. A reduced chamber pressure results in a increased diffusion constant and Thiele length. That is the reason why LPCVD refers mainly to tube reactors with closely-spaced wafers. The low pressure and the corresponding large Thiele length compared to the wafer size enables a good radial uniformity [1].

Hard-Sphere estimation for the diffusion constant

A simple estimation of the diffusion constant is based on the assumption that the molecules are hard-spheres that collide with each other corresponding the mean free path and the mean thermal velocity [1]. For the diffusion constant D follows, if only self-diffusion is considered (more details are provided i.e. by [7]),

$$D = \sqrt{\frac{\mathrm{k_B}}{\pi^3\,m}\frac{T^{3/2}}{p\,d^2}}, \tag{19}$$

where m and T are the absolute mass of the diffusing particles and their temperature, p and d are the pressure of the diffusing particle and their effective diameter, $\mathrm{k_B}$ is the Boltzmann constant.

2.5.2 Convective mass transport

Convective mass transport results from a bulk motion of the gas and requires no concentration gradient. It can be well described by fluid dynamics assuming the flow is incompressible because its velocity u is much smaller than the velocity of sound and so much smaller than the velocity of the individual molecules. The first step to describe the flow behavior might be via the Reynolds number (Re). It is a dimensionless number proportional to the ratio between inertial forces and viscous forces and therefore to the ratio between momentum convection to momentum diffusion,

$$Re = \frac{\rho\, L\, u}{\eta} = \frac{L\, u}{\nu} \quad \text{mit} \quad \eta = \nu\rho, \tag{20}$$

where ρ and u are the density and velocity of the vapor or of the gas, η and ν are the dynamic and kinematic viscosity of the gas/vapor, L describes the characteristic spatial dimension of the system.

The Reynolds number is used to analyze the flow regime, whether laminar or turbulent flow is given. At low Reynolds numbers viscous forces are dominant and consequently a laminar flow is present, which means that the gas flows in form of parallel layers that do not interfere each other. On the other hand high Reynolds numbers mean that the inertial forces are dominating, which leads to a turbulent flow with its characteristic irregularity and chaotic property changes. Therefore the turbulent regime is absolutely not suited for a modern CVD process.

Another interesting approach to understand the role of the Reynolds number for CVD system is to consider the kinematic viscosity in detail. It actually has the same dimension as the diffusion coefficient. If ν is now replaced by D in (20), we can see that the Reynolds number corresponds to the Peclet number (12). Consequently the Reynolds number can be considered as a Peclet number for the diffusion of momentum and describes - in analogy to it - the ratio between the system size and the diffusion length of the momentum [1]. If it is low, the momentum can diffuse quickly across the whole system and enables a uniform, laminar flow. But if the Reynolds number is large, the diffusion of the momentum is unsatisfactory for a uniform distribution of the momentum over the whole system and therefore velocities usually vary greatly from one part of the reactor to another corresponding to a turbulent flow.

2.5.3 Plug and Poisseuille flow

Considering a laminar flow there are two extremal cases for the flow behavior within a pipe [6]. On the one hand there is the plug flow, which is the simplest case since the gas is assumed to be not affected by viscous drag from the walls, which means a constant velocity all over the pipe.

On the other hand there is the Poiseuille flow that presupposes drag forces by the wall leading to zero velocity of the gas layer right at the walls. For reaching this state the gas has to travel a distance much longer as the entry length L_e (21). Otherwise it can be usually treated as a plug flow,

$$L_e \approx \frac{R_e}{20}. \tag{21}$$

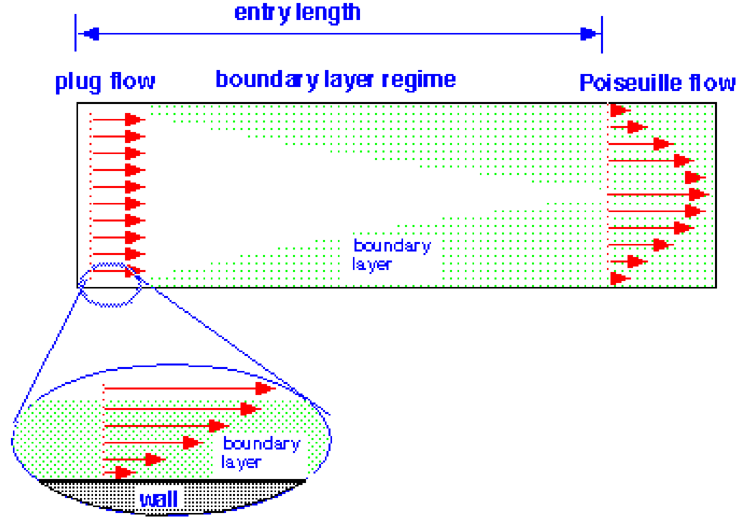


Figure 3: Schematic view of the gas flow in a pipe (after [6]). Note that at the boundary layer the gas always has zero velocity. At the beginning the gas flow is in the so called Plug regime. With exception of the boundary layer the whole cross section has the same velocity. It switches to the Poiseuille regime after the gas has traveled much longer than the entry length L_e (21), because of the dissipative forces between the gas layers.

2.5.4 Transport regimes

In case of a laminar flow the transport behavior of a reactor or at least parts of it can be generally divided into three regimes. One of them - the Thiele transport regime - has already been explained. It provides a clearly diffusion dominated transport in all directions, since the diffusion length is much longer than the relevant system length. In that case convection does not play an important role.

The second regime is called Simple Depletion [6] and can be explained also by considering a long and thin downstream region as depicted in figure 4. The diffusion in y-direction is so fast that the concentration can be considered as constant in this direction. However, in x-direction convection dominates over diffusion and is therefore determining the flow. By calculating the net flux into the marked volume because of convection and out of it because of deposition, the concentration gradient in x-direction can be estimated analogously to the Thiele transport regime. Again it is assumed that deposition only occurs on the bottom surface.

The net flux into the marked volume because of convection is

$$\begin{aligned}\Phi_{\text{conv}} &= \Phi_{\text{conv}}^{in} - \Phi_{\text{conv}}^{out} = ZH(J_{\text{conv}}^{in} - J_{\text{conv}}^{out}) = \\ &= ZH\left(U\,n(x) - U\,n(x+\mathrm{d}x)\right) = ZHU\left(\frac{n(x) - U\,n(x+\mathrm{d}x)}{\mathrm{d}x}\,\mathrm{d}x\right) = -ZHU\,\frac{\mathrm{d}n}{\mathrm{d}x}\,\mathrm{d}x.\end{aligned} \tag{22}$$

Now this flux can be set equal with the deposition induced flux Φ_{dep} (15) out of the marked volume assuming steady state,

$$\Phi_{\text{conv}} = \Phi_{\text{dep}} \Rightarrow -ZHU\,\frac{\mathrm{d}n}{\mathrm{d}x}\,\mathrm{d}x = Z\,\mathrm{d}x\,K_s\,n \Rightarrow \frac{\mathrm{d}n}{\mathrm{d}x} + \frac{K_s}{HU} = 0, \tag{23}$$

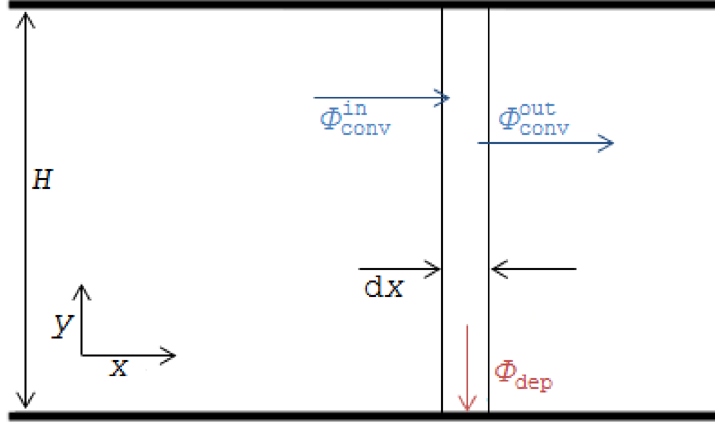


Figure 4: Schematic view of the simple depletion transport regime. The precursor concentration n is constant in the infinitesimal volume with the thickness dx.

with the solution

$$n(x) = n(0)\, e^{-x/L_{sd}}, \quad L_{sd} = \frac{HU}{K_s}. \tag{24}$$

Concerning the zero-dimensional reactor model, the ratio H/K_s can be identified as the consumption time t_{cons} according to (3). The depletion length is then the product of consumption time and gas velocity,

$$L_{sd} = t_{\text{cons}}\, U, \quad t_{\text{cons}} = \frac{H}{K_s}. \tag{25}$$

In the third flow regime also convection dominates in the x-direction like in the simple depletion regime with the difference that diffusion in y-direction is slow compared to the deposition rate [6]. This leads to a gradient in y-direction with a concentration of zero at the tube boundaries. The problem is depicted in figure 5.

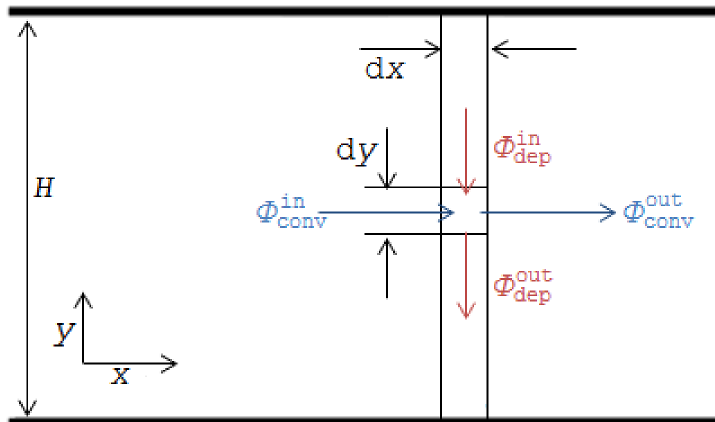


Figure 5: Schematic view of the transport regime, where convection dominates in the x-direction and diffusion dominates in the y-direction. Since it is assumed that diffusion in y-direction is slow against the deposition rate, the axial concentration at the borders can be approximated as zero.

Accordingly, the flux in and out of the marked volume is composed of the net flux of convection in x-direction,

$$\Phi_{\text{conv}} = Z\,\mathrm{d}y\,(J^{in}_{\text{conv}} - J^{out}_{\text{conv}}) = -Z\,\mathrm{d}y\,U\frac{\mathrm{d}n}{\mathrm{d}x}\mathrm{d}x, \tag{26}$$

and the net flux of diffusion in y-direction,

$$\Phi_{\text{diff}} = Z\,\mathrm{d}x\,(J^{in}_{\text{diff}} - J^{out}_{\text{diff}}) = -Z\,\mathrm{d}x\,D\,\frac{\mathrm{d}^2 n}{\mathrm{d}y^2}\,\mathrm{d}y. \tag{27}$$

By equating them this gives the partial differential equation

$$\Phi_{\text{conv}} = \Phi_{\text{diff}} \Rightarrow Z\,\mathrm{d}y\,U\,\frac{\mathrm{d}n}{\mathrm{d}x}\,\mathrm{d}x = Z\,\mathrm{d}x\,D\,\frac{\mathrm{d}^2 n}{\mathrm{d}y^2}\,\mathrm{d}y \Rightarrow \frac{\mathrm{d}^2 n}{\mathrm{d}y^2} - \frac{U}{D}\,\frac{\mathrm{d}n}{\mathrm{d}x} = 0, \tag{28}$$

that actually has the form of the one dimensional heat conduction equation

$$\frac{\partial^2 n}{\partial y^2} - a\,\frac{\partial n}{\partial x} = 0, \quad 0 < x < H, t > 0, \tag{29}$$

with the boundary conditions

$$u(0,x) = 0, \quad u(2H,x) = 0, \tag{30}$$

where y were originally the single spatial dimension, x the chronological dimension, a the thermal diffusivity and n the temperature. Its general solution is

$$n(y,x) = \sum_{m=1}^{\infty} n_m\,\sin\left(\frac{m\pi}{H}y\right)\,e^{-\left(\frac{m^2\pi^2}{H^2 a}x\right)}. \tag{31}$$

The corresponding solution for the flow problem is then

$$n(y,x) = \sum_{m=1}^{\infty} n_m\,\sin\left(m\pi\frac{y}{H}\right)\,e^{-\left(m^2\pi^2\frac{x}{L_{rsr}}\right)}, \quad L_{rsr} = \frac{H^2 U}{D}. \tag{32}$$

In order to sum up the three flow regimes their conditions are listed below in combination with the corresponding precursor concentration functions:

- **Thiele transport:** *Diffusion dominated in all spatial dimensions*
  Conditions: $U << D/L$
  $n(x) = n(0)\,e^{x/L_{th}}, \quad L_{th} = \sqrt{\frac{HD}{K_s}}$
- **Simple depletion:** *Convection is dominant in x-direction, while diffusion provides a constant concentration in y-direction*
  Conditions: $U >> D/L, \quad K_s << D/H$
  $n(x) = n(0)\,e^{-x/L_{sd}}, \quad L_{sd} = \frac{HU}{K_s}$
- **Rapid surface reaction:** *Convection is dominant in x-direction, while diffusion is determinant for the y-direction. Rapid surface reactions lead to a concentration of zero at the surface which causes a gradient in y-direction*
  Conditions: $U >> D/L, \quad K_s >> D/H$
  $n(y,x) = \sum_{m=1}^{\infty} n_m\,\sin\left(m\pi\frac{y}{H}\right)\,e^{-\left(m^2\pi^2\frac{x}{L_{rsr}}\right)}, \quad L_{rsr} = \frac{H^2 U}{D}$

These three simple solutions for different reactor conditions can also be used for describing more complex geometries. A simple combination of them gives at least a clue what is going on in the reactor concerning mass transport.

2.6 Reaction chemistry

The main question concerning the reaction chemistry of a CVD process may be under which conditions the desired solid layer is deposited. Basically there are two approaches. The first one goes over simple thermodynamics while the second one uses a detailed kinetic analysis of all included reactions, which is generally a huge effort because of the usually large amount of possible reactions. However, a kinetic analysis gives also information about the reaction rates in comparison to a simple thermodynamical analysis which only allows to answer the question whether the deposition takes place or not and what the actual equilibrium state of the system is. The thermodynamical analysis does not provide information about the reaction speed, or whether the equilibrium is ever reached in an appropriate time. [6]

2.6.1 Thermodynamic approach

Basically the direction of the reaction as well as the equilibrium state of the system provide the information whether the desired reaction happens under the considered conditions. The direction of the reaction as well the equilibrium state can be estimated by minimizing the Gibbs free energy G. If it is negative the reactions run in the right direction and lead to the desired products. The change of the free energy is defined as

$$\Delta G = \Delta G^{\circ} + RT \ln \prod_i a_i^{\nu_i}, \tag{33}$$

where a_i stands for the activity of the species i. In case of gases it is

$$a_i = \frac{p_i}{p} = x_i, \tag{34}$$

where p_i and x_i are the partial pressure and the mole fraction of the species i, respectively. ΔG° is the overall standard free energy for the CVD process,

$$\Delta G^{\circ} = \sum_i \nu_i \, \Delta G_i^{\circ}, \tag{35}$$

at a standard temperature of $25\,^{\circ}\mathrm{C}$ and a standard pressure of $1\,\mathrm{atm}$. ν_i represents the stochiometric coefficient of species i in the CVD reaction. It is negative for reactants and positive for products. ΔG_i° is the standard free energy for the formation of the species i at standard temperature and pressure.

Since a minimization of the Gibbs free energy is equivalent to $\Delta G = 0$, the so called equilibrium constant K can be determined,

$$\Delta G^{\circ} = -\mathrm{R}T \ln K. \tag{36}$$

K represents a ratio of activities at which a net change of the involved reactants is impossible for energetic reasons. Assuming a reaction of the form

$$\alpha A + \beta B + \gamma C + ... \rightleftharpoons \rho R + \sigma S + \tau T + ..., \tag{37}$$

where the equilibrium constant K is

$$K = \prod_i x_i^{\nu_i} = \frac{x_R^{\rho} x_S^{\sigma} x_T^{\tau}}{x_A^{\alpha} x_B^{\beta} x_C^{\gamma}} = e^{-\frac{\Delta G^{\circ}}{\mathrm{R}T}}. \tag{38}$$

For a precise thermodynamical analysis it is necessary to consider all involved reactions, which is usually a bunch for a typical CVD process. The set of concentrations that actually satisfy all the obtained reaction equations has to be determined by optimization. However, this is even more difficult than a detailed kinetic analysis because there is usually few information existing about the bulk of standard Gibbs free energies of the involved reactions [6].

Nonetheless, this method can be used to classify the reaction by only taking into account the overall net reaction of the considered CVD process. If the change of the corresponding standard Gibbs free energy is small the process is reversible in principle. Depending on the initial conditions the process can be adjusted between deposition or etching. If the change of the Gibbs free energy is strongly negative, the process is virtually irreversible and only deposition takes place. [6]

Table 1: Listing of the standard gibbs free energy of a few common CVD processes [6]. ΔG° is the standard Gibbs free energy

Process	ΔG° [kJ/mol]
$SiH_4 + 2O_2 \rightarrow SiO_2 + 2H_2O$	-1307
$SiH_4 + O_2 \rightarrow SiO_2 + 2H_2$	-850
$3SiH_4 + 4NH_3 \rightarrow Si_3N_4 + 12H_2$	-742
$SiH_4 \rightarrow Si + 2H_2$	-57
$TiCl_4 + 2NH_3 \rightarrow TiN + 4HCl + H_2$	92
$TiCl_4 + 2H_2 \rightarrow Ti + 4HCl$	287

According to table 1 the deposition of SiO_2 from SiH_4 and O_2 including both possible ways is highly exergonic even at room temperature so that the process is virtually irreversible. In comparison to that the transformation of $TiCl_4$ to TiN or Ti is slightly to moderately endergonic under standard conditions so that the reaction has to be forced with high temperatures of about $1000\,^{\circ}C$.

The top layer of the substrate also affects the free energy of the deposition process, which is not included in table 1. Although this effect is usually fairly small it has a considerable influence on the deposition if the process is driven near the equilibrium state, which means that the change of the Gibbs free energy is almost zero. In that case the influence of the top layer on G decides whether deposition or etching takes place. An example is the selective deposition of bulk silicon Si from a SiH_2Cl_2, HCl mixture [6]. Under appropriate conditions net deposition of Si only takes place on Si-surfaces but not on SiO_2-surfaces.

2.6.2 Kinetics

A kinetic approach provides a deep insight into the rate of all involved reactions. A reaction with a high negative Gibbs free energy does not automatically implicate that it

is taking place rapidly. It can be so slow that it plays no role within the overall process. Or it can - according to the principle of the weakest member of the chain - rather be rate determining the overall reaction if no alternative path is leading around this reaction.

Rate expressions

Despite of all the advantages of a kinetic analysis, a comprising knowledge or investigation about the mostly large number of reaction as well as their properties is required. Already at a simple reaction of the form $A + B \rightarrow AB$ the rate is not inevitably equal to $kc_A c_B$ if subreactions take place.

However, assuming a direct association reaction of the form $A + B \rightarrow AB$, the reaction rate r has a linear dependence of the concentrations of the involved species,

$$r = \frac{\mathrm{d}c_{AB}}{\mathrm{d}t} = k\, c_A c_B, \tag{39}$$

where k is the rate coefficient that can fully be described by the modified Arrhenius equation [1] in most cases,

$$k = A\, T^B\, e^{\frac{E_a}{\mathrm{k_B}T}}, \tag{40}$$

where E_a is the activation energy and A and B are reaction depended coefficients. Especially the activation energy does have a large influence on the feasibility of the corresponding reaction under certain conditions, because already small changes do have a considerable impact. Thus, the strong temperature dependence of the exponential term frequently covers over the other temperature term T^B. This is the reason why often only the original Arrhenius equation

$$k = A_1\, e^{\frac{E_a}{\mathrm{k_B}T}}, \tag{41}$$

is used, where AT^B is approximated by the pre-exponential factor A_1.

Transition states

Since an association reaction ($\mathrm{A} + \mathrm{B} \rightarrow \mathrm{AB}^{**}$) automatically leads to an activated complex AB^{**} that dissociates again after a characteristic time, a third partner is required during or after the association reaction to take energy, allowing to get the compound more stable. Another but rather unlikely option is a radiative decay that leads to a deactivation of the complex. So in a CVD process the overall concentration of such a complex AB often depends on such a third partner. At lower pressure this means a highly pressure dependent concentration while at high pressures the pressure dependence is negligible. [1]

Assuming that the transition state shows a quasi-equilibrium with the reactants, the concentration is independent of the exact reaction path,

$$c^{**} = k^{**}\, c_A c_B, \tag{42}$$

where k^{**} is the equilibrium constant which is

$$k^{**} = \frac{\mathrm{R}T}{p_0}\, e^{-\frac{\Delta G}{\mathrm{R}T}} = \frac{\mathrm{R}T}{p_0}\, e^{-\frac{\Delta H}{\mathrm{R}T}}\, e^{\frac{\Delta S}{\mathrm{R}}}. \tag{43}$$

The change of enthalpy ΔH is determined by the barrier height while the entropy part takes into account the amount of identical transition states. In order to obtain the rate of that association reaction, c^{**} has to be multiplied by the frequency of thermal vibrations kT/h. [1]

Lindemann-Hinshelwood mechanism
The Lindemann-Hinshelwood mechanism splits up stepwise reactions into multiple elementary steps each of them possessing an own rate constant. Usually the slowest step would be rate-determining, but in a steady state the concentrations of the intermediates are adjusted so that their net change is zero. Therefore in a steady state all elementary steps are contributing to the rate. Considering the excitation of species A by a particle M the involved reactions can be:

- Inelastic collision between the species A and the particle M that leads to the excitation of A. The corresponding rate is:
  $\left.\frac{\mathrm{d}c_A^*}{\mathrm{d}t}\right|_{ex} = k_{ex}\, c_A c_M$
- Superelastic collision between the excited species A^* and the particle M that leads to the de-excitation of A. The corresponding rate is:
  $\left.\frac{\mathrm{d}c_A^*}{\mathrm{d}t}\right|_{dx} = -k_{dx}\, c_A^* c_M$
- Radiative decay or decomposition into the products of A. Both processes lead to a de-excitation of A^* and are dependent to its concentration:
  $\left.\frac{\mathrm{d}c_A^*}{\mathrm{d}t}\right|_{de} = -k_{de}\, c_A^*$

The total change of c_A^* is given by the sum of the three mentioned components. Assuming steady state it has to be zero,

$$\frac{\mathrm{d}c_A^*}{\mathrm{d}t} = k_{ex}\, c_A c_M - k_{dx} c_A^*\, c_M - k_{de}\, c_A^* \overset{!}{=} 0. \tag{44}$$

Consequently the concentration c_A^* is

$$c_A^* = \frac{k_{ex}\, c_A c_M}{d_{dx}\, c_M + d_{de}}. \tag{45}$$

If the pressure is sufficiently low, the superelastic collision term can be neglected with respect to the third term so that the concentration c_A^* is of second-order, inverse of pressure, since the concentration corresponds to the density and therefore to the inverse of the pressure. Vice versa if the pressure is high enough the superelastic collision term covers up the third term and leads to a concentration c_A^* that is of first-order in the inverse of pressure.

Chain reactions induced by radicals
The activation energy is high for many reactions in CVD processes, so that the reaction rate is accordingly low. In order to bypass that problem and to increase the rate, catalysis mechanisms are needed. In detail highly reactive species decrease the activation energy by pasting intermediate steps that run with a much higher rate. In general radicals play this role. In CVD processes they usually arise from decomposition of the precursors. In fact, they are included in the majority of the reactions.

Every species with an odd number of valence electrons is a radical. On the other hand all non-radicals must have an even number of valence electrons. So one special property of radical reactions is that a radical (odd number) that reacts with a non-radical (even number) results again in a radical (odd + even = odd) that itself can

participate in another reaction to create again a radical. This chain reaction only ends when two radicals react with each other so that a non-radical arises. However, this is a rather rare event, especially at low pressure, since a radical-radical reaction usually needs a third partner that absorbs some of the de-excitation energy that eventually enables a stable product. In order to start such a chain reaction, an initial event has to occur that decomposes a non-radical into at least one radical.

What has not been mentioned until now is that not all species with an even number of valence electrons have to be a non-radical. Actually every species with an unpaired electron is a radical. For example atomic oxygen O, which has an even number of valence electrons but is nevertheless a radical, since it has two unpaired electrons. Thus, it is often possible that in typical association reactions of a radical and a non-radical, two or more radicals can arise, which is called a branching radical chain [1]. This leads to an increasing number of radicals with each iteration. If this happens uncontrolled an explosion takes place.

A very well-known example is the oxyhydrogen explosion where hydrogen (H_2) and oxygen (O_2) react which each other, initiated by an dissociated hydrogen molecule that leads to two hydrogen radicals H. The subsequent and most important reactions are

$$H\cdot + O_2 \rightarrow OH\cdot + \cdot O\cdot \quad (46)$$

$$\cdot O\cdot + H_2 \rightarrow OH\cdot + H\cdot \quad (47)$$

$$OH\cdot + H_2 \rightarrow H_2O + H\cdot. \quad (48)$$

As it can be seen the overall reaction spreads highly into branches leading to a increasing number of radicals. The final stable product is water with an overall released energy of $-286\,\mathrm{kJ/mol}$ (assuming standard conditions).

Another example, which is much more interesting for this work is the explosive behavior of SiH_4 (silane) in combination with O_2. In semiconductor manufacturing this mixture is used in controlled CVD processes to produce and deposit silicon dioxide (SiO_2), where the large number of arising radicals, especially from SiH_4 takes a major role in the process chemistry. Uncontrolled, under standard conditions SiH_4 is highly pyrophoric and consequently explosive, since oxygen (O_2) and usually water (H_2O) are present in air. Both radicals OH and H which arise of spontaneous decomposition of water contribute to a branching chain reaction in combination with silane,

$$OH\cdot + SiH_4 \rightarrow SiH_3\cdot + H_2O \quad (49)$$

$$H\cdot + SiH_4 \rightarrow SiH_3\cdot + H_2. \quad (50)$$

The direct reaction of OH or H with SiH_4 does not lead to an increasing number of radicals. However, the following reaction of SiH_3 with O_2 does so,

$$SiH_3\cdot + O_2 \rightarrow SiH_3O\cdot + O\cdot \quad (51)$$

$$O\cdot + SiH_4 \rightarrow SiH_3\cdot + OH\cdot \quad (52)$$

$$\ldots. \quad (53)$$

The generated radicals can react again with SiH_4 or SiH_3 and other intermediates producing more and more radicals leading to the formation of SiO_2, eventually. The variety of paths at this process is enormous and will be discussed later. Overall, this is the reason why a silane-air mixture is so dangerous.

2.6.3 Surface reaction kinetics

Only a little fraction of the deposited film is produced by direct formation of the deposition material within the gas phase with subsequent deposition. Consequently the major other part is caused by several surface reactions fed by adsorbing compounds of the gas phase. In detail dangling bonds, which are unsaturated valences on immobilized atoms that are sitting on the surface and leading to high energies, react mainly with radicals in multiple steps to a final saturated compound, which minimizes the Gibbs free energy. Although reactions of dangling bonds with non-radicals are sometimes possible, they are usually of minor influence since their sticking coefficient is much lower than that of chemically similar radicals.

In general adsorption, which means an adhesion between gas species and the surface, can be classified into physisorption and chemisorption. Physisorption is a process, where the attractive forces are provided by van der Waals forces. Although its binding energy is very weak it plays an important role for CVD processes. Moreover it is often the preliminary stage of chemisorption, which means covalent bonds between gas species and the surface. These covalent bonds, which provide a much stronger binding energy, lead to immobile surface compounds such as the mentioned dangling bonds.

Langmuir isotherms

Langmuir Isotherms describe the surface coverage of a certain species with respect to its gas phase concentration and its partial pressure, respectively. It is assumed that adsorption is limited by one monolayer and that all sorption places are equivalent. Further it is assumed that the surface of the sorption places is uniform and that there are no interactions between them.

Of course, the sorption rate of a certain gas phase species A is influenced not only by the adsorbing species but also the desorbing species. The adsorption rate of species A is proportional to the gas phase concentration n_A which is in turn proportional the its partial pressure p_A. It is also proportional to the free sorption places $(1-\theta)$ and to an adsorption constant k^A_{ads}. On the other hand the desorption rate is only proportional to the proportion of sorption places θ_A occupied by species A and to a desorption constant k^A_{des}. Therefore, it follows for the sorption rate of species A

$$\frac{\mathrm{d}N^A_{srp}}{\mathrm{d}t} = k^A_{ads}\, p_A\, (1-\theta)\, N - k^A_{des}\, \theta_A N, \tag{54}$$

where N indicates the total number of sorption places. Considering steady state, $\frac{\mathrm{d}N^A_{srp}}{\mathrm{d}t}$ has to be zero. It follows for θ_A, assuming a gas phase with only that one species A $(\theta = \theta_A)$,

$$\theta_A = \frac{k^A_{ads}\, p_A}{k^A_{des} + k^A_{ads}\, p_A} = \frac{K_A\, p_A}{1 + K_A\, p_A}, \tag{55}$$

where $K_A = k^A_{ads}/k^A_{des}$ is introduced as sorption coefficient. The coverage θ_A is linear at a low partial pressure p_A and is saturated at high partial pressures (figure 6). The Langmuir isotherm can be of course extended to multiple gas phase species that compete for the sorption places. For two gas phase species these are

$$\theta_A = \frac{K_A\, p_A}{1 + K_A\, p_A + K_B\, p_B}, \quad \theta_B = \frac{K_B\, p_B}{1 + K_A\, p_A + K_B\, p_B}. \tag{56}$$

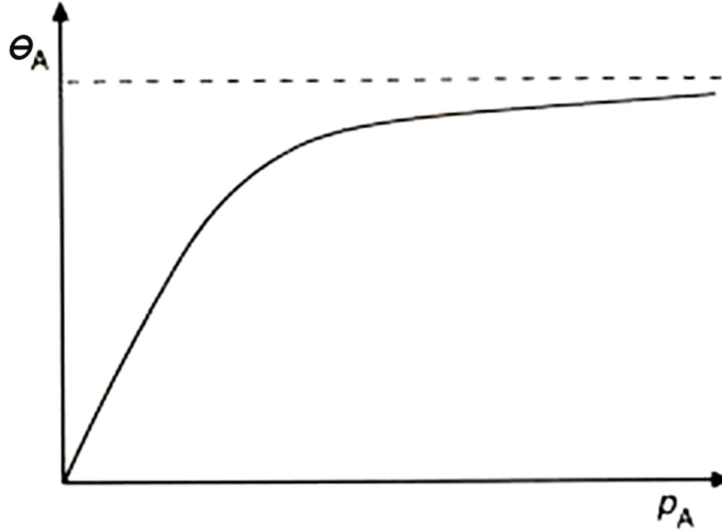


Figure 6: General form of a Langmuir-isotherm (after [8]). The partial pressure of all species with exception of species A is kept constant.

Langmuir-Hinshelwood mechanism

Let us consider a gas phase with two species A and B that react to species C when they are both adsorbed at the surface. Species C becomes then part of the surface and produces itself a free sorption place. Overall, the total number of sorption places then does not change by A and B adsorbing and reacting with each other. The reaction rate r_{AB} is proportional to the coverage of A and B and to a reaction coefficient k, assuming that the coverage is of the form of Langmuir isotherms (56),

$$r_{AB} = k\,\theta_A \theta_B = \frac{k\,K_A\,p_A\,K_B\,p_B}{(1 + K_A\,p_A + K_B\,p_B)^2}. \tag{57}$$

Equation (57) shows the dependence of the reaction rate r_{AB} with respect to the ratio of the partial pressures of A and B (p_A, p_B). If species C is already the stable final product of the layer, the reaction rate r_{AB} can be equated with the deposition rate.

As it is depicted in figure 7 r_{AB} has its maximum at a specific ratio $\frac{p_A}{p_B}|_{opt}$. Assuming that p_B is fixed and considering now the situation that p_A is varied so that $\frac{p_A}{p_B}$ is much smaller than $\frac{p_A}{p_B}|_{opt}$, the reaction rate is proportional to p_A. If $\frac{p_A}{p_B}$ is much higher than $\frac{p_A}{p_B}|_{opt}$ the reaction rate is indirectly proportional to p_A. This specific behavior results from the competition of species A and B for the free sorption places. If one species is present too strongly it saturates the surface by occupying a large part of the potential sorption places leading to a reduced reaction rate.

Note that this sort of surface reaction mechanism is rarely in common CVD processes, since preferably immobile dangling bonds are reacting with just adsorbed gas phase species in multiple steps to the stable final product. Though, this is a possible surface reaction that can probably occur.

Eley-Rideal mechanism

This mechanism assumes that only one component (species A) is adsorbed by the surface. The other component (B) participates in the reaction directly from the gas phase.

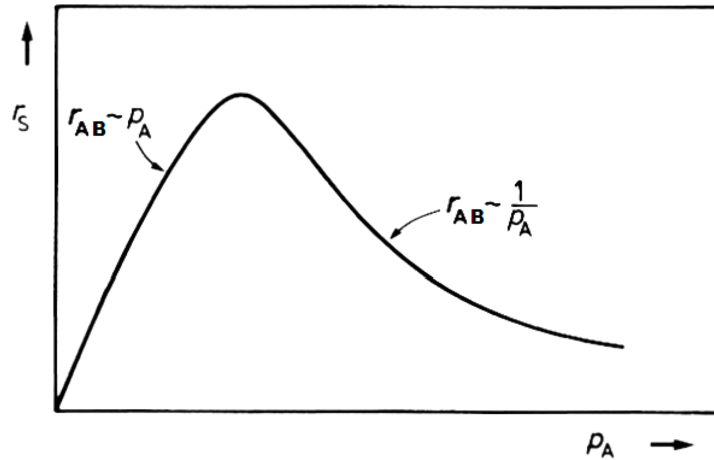


Figure 7: Langmuir-Hinshelwood mechanism (after Kasper [9]). Reaction rate r_{AB} as function of the partial pressure p_A of species A. It reaches its maximum at an optimum partial pressure ratio of species A and B.

Consequently the reaction rate r_{AB} is proportional to the coverage of species A (θA). However, concerning species B it is only proportional to its partial pressure p_B,

$$r_{AB} = k\,\theta_A p_B = k\,\frac{K_A\,p_A}{1 + K_A\,p_A}\,p_B, \tag{58}$$

where k indicates a specific reaction constant. Furtheron, in the second step of equation (58) θ_A is assumed have the form of Langmuir isotherms. So the reaction rate does not show a maximum in comparison to the Langmuir-mechanism. [9]

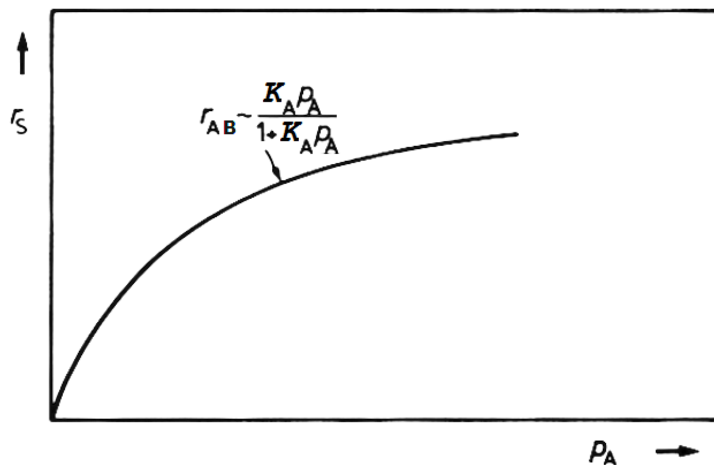


Figure 8: Eley-rideal mechanism (after Kasper [9]). Reaction rate r_{AB} as function of the partial pressure p_A of species A . While only species A is adsorbed by the surface, species B participates directly from the gas phase.

2.6.4 Film conformality

An important question in chemical vapor deposition of any films is the conformality of that specific film, which is treated in the work of Zuraw and Dobkin [1]. Conformality

actually means the quality of gap filling and the step coverage, respectively. A bad conformality resulting in free spaces in between two closely spaced walls does have a big influence on many properties of the film. Especially in semiconductor manufacturing the electrical properties are affected. In order to answer the conformality question it seems appropriate to take a closer look on the probability of adsorption, desorption and the reaction.

An incoming precursor faces three options. While it can firstly be just reflected by the wall, it can secondly be adsorbed followed by desorption. However, the third option is that the precursor adsorbs and subsequently reacts with the surface or with another adsorbed precursor. The probability of this third option is, however, now crucial for the film conformality.

If this so called reactive sticking coefficient RSC is unity, the corresponding precursor only hits the surface once. Since precursors usually have a steep angle of incidence, the horizontal surfaces are more exposed and are growing faster. Furthermore, the film is growing faster at edges which eventually leads to a concrescence of the upper part by leaving behind free spaces at the lower part of gaps with a higher aspect ratio (compare figure 18). If the RSC is close to zero, the precursor are reflected or desorbed multiple times which enable them to reach every point of the 3d-structure from every direction leading to a constant rate of reactive sticking at every point of the 3d-structure. This enables a very uniform growing of the film independent of the orientation and location of the surface, which results in a rather conform film.

In order to estimate the conformality, it is sufficient to consider only the RSCs of the reactants of the dominant surface reactions. For example, if preferably radicals are involved as reactants with certain dangling bonds, the RSC is consistently high leading to a rather mediocre conformality.

For a judgment of the degree of conformality the total change ΔH of the CVD process can surprisingly be considered as a rough estimation. With a higher ΔH the conformality is usually better. For example, ΔH of silane with oxygen reacting to silicon dioxide is strongly negative resulting in a mediocre conformality. On the other hand ΔH of silane reacting to bulk silicon is very weakly negative, which leads to an excellent conformality. An obvious explanation for this trend is that higher energies tend to produce more reactive intermediates, which therefore have a higher RSC.

3 Plasma enhanced CVD

Plasma Enhanced CVD (PECVD) generally goes along with a plasma state where 'hot' electrons exist in a much cooler ensemble of atoms and molecules. This is accomplished by a typical low pressure of about 0,5-20 Pa. It has to be ensured that the external energy is mainly coupled to the electrons. Indeed, since electrons transfer only a very small amount of its energy via elastic collisions to much heavier atoms or molecules, it would need several thousand collisions for every single electron to reach a thermalisation between electrons and atoms and molecules. This is mostly prevented by a short residence time. In that case the high energetic (hot) electrons are mainly responsible for the decomposition processes of the precursors and intermediates.

Thus, a deposition at much lower temperature is enabled by PECVD compared to many thermal CVD processes. This is one of the main advantages of PECVD, since many substrates as i.e. wafers are temperature sensitive. Secondly, PECVD also enables conditionally a bombardment of the substrate by high energetic ions. Especially capacitively or partly capacitively driven plasmas provide such a behavior. An ion bombardment increases the density of the deposited film significantly and makes it more compact. These enhanced properties can become very important in applications where the thickness of the film is limited (i.e. for geometrically reasons), but has to fulfill high demands. For example isolating SiO_2 layers between metal lines have to be quite dense and uniform in order to isolate sufficiently. For this purpose the HDP CVD process is used, which is a PECVD process that includes an excessive ion bombardment that leads to sputtering.

The third big advantage is its easy cleaning in comparison to thermal CVD processes. Considering again the HDP CVD process it is sufficient to feed the reactor with a fluorine-containing gas like NF_3. By running a plasma, the fluorine-containing molecules get decomposed into highly reactive fluorine radicals that combine subsequently with the immobile SiO_2 into mobile species, which are removed by the pumps, eventually. This is very important in practical applications, since deposition not only takes place on the substrate but also on all other parts of the reactor chamber. Although the corresponding thickness of the deposited material is usually smaller than on the substrate due to the temperature differences, it has to be removed in order to keep the chamber conditions constant. Moreover, the deposited material on the chamber walls would produce more and more particles because of i.e. temperature induced stress.

On the other hand the disadvantages of PECVD are potential risks of plasma damage on the substrate, which is discussed later on. The complexity of the reactor is increased in comparison to a conventional thermal CVD reactor. Among other reasons because of the larger amount of power supplies, where each of them has to be provided by an appropriate matching network [1]. Or because of the powered electrodes, which have to be electrically isolated by durable materials, as well as substrate heaters that also have to be isolated, etc.

3.1 Ambipolar diffusion

Electrons and ions are moved analogous to neutrals due to diffusion caused by a nonzero density gradient of the corresponding particles. But additionally to neutrals there

also exists a drift force for charged particles because the electric field that occurs with the non-zero density gradient of the charged particles. The flux of the ions and electrons is

$$J_i = \mu_i n_i E - D_i \frac{\partial n_i}{\partial x}, \quad J_e = -\mu_e n_e E - D_e \frac{\partial n_e}{\partial x}, \tag{59}$$

assuming one spatial dimension. An additional external electric field created by the PECVD reactor will be excluded in this chapter. It just would create an additional force that would increase the movement of the charged particles. Here only the movement of charged particles caused by internal forces is of interest. [1]

Since the difference between ion and electron density is in general very small at such plasma conditions, it can be approximated that they are equal, $n_i = n_e = n$. Analogously the two fluxes can be assumed to be equal, $J_i = J_e = J$. Thus, it follows for the electric field E from the two equations (59)

$$\mu_i n\, E - D_i \frac{\partial n}{\partial x} = -\mu_e n\, E - D_e \frac{\partial n}{\partial x} \Longrightarrow E = \frac{D_i - D_e}{\mu_i + \mu_e} \frac{\frac{\partial n}{\partial x}}{n}. \tag{60}$$

By inserting this electric field in the general form of the ion flux (59) one obtains (of course the same result would be obtained by using the electrons flux),

$$J = J_i = \mu_i n \left(\frac{D_i - D_e}{\mu_i + \mu_e} \frac{\frac{\partial n}{\partial x}}{n} \right) - D_i \frac{\partial n}{\partial x} = -\frac{\mu_i D_e + \mu_e D_i}{\mu_i + \mu_e} \frac{\partial n}{\partial x}. \tag{61}$$

The term before the first derivative of the charge carrier density n can be identified as a new diffusion coefficient, the so-called ambipolar diffusion coefficient D_a,

$$D_a = \frac{\mu_i D_e + \mu_e D_i}{\mu_i + \mu_e}. \tag{62}$$

This ambipolar diffusion coefficient can be simplified, because the mobility of the ions can be assumed to be very small compared to the mobility of the electrons ($\mu_i << \mu_e$),

$$D_a \approx D_i + \frac{\mu_i}{\mu_e} D_e. \tag{63}$$

Now the Einstein-Smoluchowski relation that determines the ratio between diffusion and mobility,

$$\frac{D}{\mu} = \frac{\mathrm{k_B} T}{q}, \tag{64}$$

where q is charge of the corresponding charge carrier, can be used to further simplification of the ambipolar diffusion coefficient,

$$D_a = D_i \left(1 + \frac{T_e}{T_i} \right). \tag{65}$$

Since the electron temperature T_e is much higher than the ion temperature T_i under such plasma conditions, the ambipolar diffusion coefficient D_a is approximately the diffusion coefficient of the ions D_i times the ratio of T_e and T_i. Since at typical plasma conditions in a PECVD reactor the electron temperature is 50-100 times larger, this is actually the typical ratio of D_a and D_i. So even at conditions in a plasma reactor where convective mass transport dominates for neutrals, it can be assumed that diffusion dominates for electrons and ions. [1]

3.2 Capacitively driven plasmas

3.2.1 Introduction

A capacitively driven plasma arrangement is often used in PECVD applications because it is not only simpler to build than other excitation methods but also very functional. Thus, for example it provides as the only one a simultaneously ion bombardment. However, there are limitations to be considered. Especially the electron density, which is so important for decomposition of precursors, is limited somehow, since a further increase of the input power would have mainly an impact to the ion bombardment and hot electron creation [1]. And this would eventually lead to sputtering of the chamber walls with a subsequent contamination of the substrate [1]. But nevertheless capacitively driven plasma arrangements are also used when higher electron densities are required. In this cases in combination with other excitation techniques, where the capacitively driven part is mainly responsible for the ion bombardment.

The electron density in an exclusively capacitively driven plasma n_e is typically about 10^8 to $10^{10}\,\mathrm{cm}^{-3}$ [1]. Assuming an electropositive plasma, where most ions are positive, the plasma can be considered to have a charge balance so that the ion density n_i is equal to the electron density. This is the case for most CVD applications, mainly because of the properties of the precursors. Electronegative plasmas, which exist preferably of cations and the same amount of anions but of much less electrons, occur rather in plasma etching applications. But this is not the topic here, so from now on always an electropositive plasma is assumed.

In order to excite a plasma capacitively, typically an RF diode is used. Between two electrodes a voltage with a frequency up to some MHz (typically 13,56 MHz) is applied. One electrode is usually connected to the power supply while the other one is grounded. Since this configuration corresponds to a capacitor it is called a capacitively driven plasma. As a consequence of the electric field between the electrodes, atoms and molecules get ionized producing secondary electrons or just get decomposed or excited, leading to light emission.

3.2.2 Construction and principle

Since the electrons have a much higher mobility as the massive ions, they impinge the electrodes far more often so that there a negative charge is quickly established, leading to a time-averaged negative potential at the electrodes while the plasma bulk is always the most positive part in the system (figure 10). The region where the potential differences arise between the electrodes an the plasma bulk is called sheath region and shows typically a size of 0,5-2,5 mm. In this region both the electron and the ion density drop strongly in comparison to the plasma bulk. A drain off of the negative charges at the grounded electrode is usually denied by a capacitor (figure 9). So the RF (AC) current can flow, while the DC field is preserved.

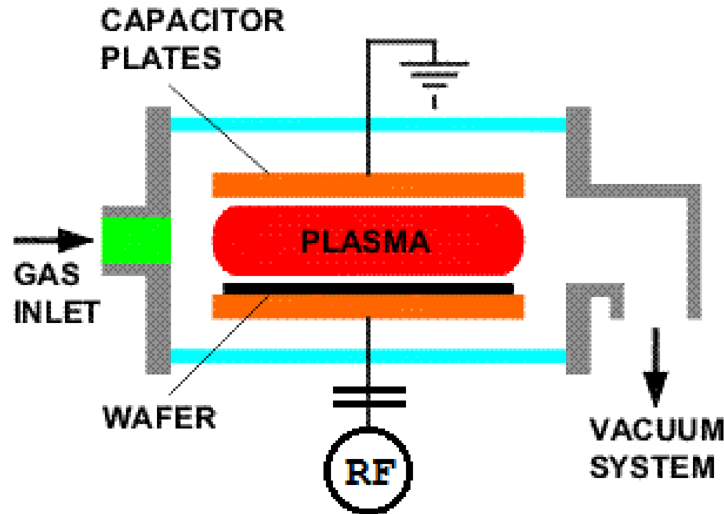


Figure 9: Schematic depiction of a capacitively coupled plasma reactor (after [10]). Note that between the bright plasma (marked red) and the electrodes (orange) spaces appear, which correspond to the sheaths. Thus, the red cloud corresponds to the plasma bulk. The additional capacitor in the circuit prevents a DC current. Consequently only AC current generated by the RF generator is possible.

As a consequence of the decrease of the potential at the beginning of the sheath regions, ions that are located there are accelerated towards the electrodes and impinge them, eventually. Thereby secondary electrons can be released, which are subsequently accelerated into the plasma. Note that this effect is amplified the smaller the electrode is, since this has a direct influence on the potential drop in the sheath regions. The voltage drop between the two electrodes because of this effect is called auto-bias and can be calculated as a function of the ratio of the electrode areas [12],

$$\frac{U_1}{U_2} \approx \left(\frac{A_2}{A_1}\right)^2 . \tag{66}$$

In this empirically found relation [1] U_1 and U_2 are the time-averaged sheath voltages of the two electrodes. $A_{1,2}$ are the corresponding areas. If the driven electrode is smaller than the grounded one, the auto-bias will be negative. Consequently, the potential drop at the driven electrode is larger (compare figure 11) and the ions are accelerated more than at the other electrode, leading to a more efficient ion bombardment. In fact this arrangement is very common in industrial applications [12].

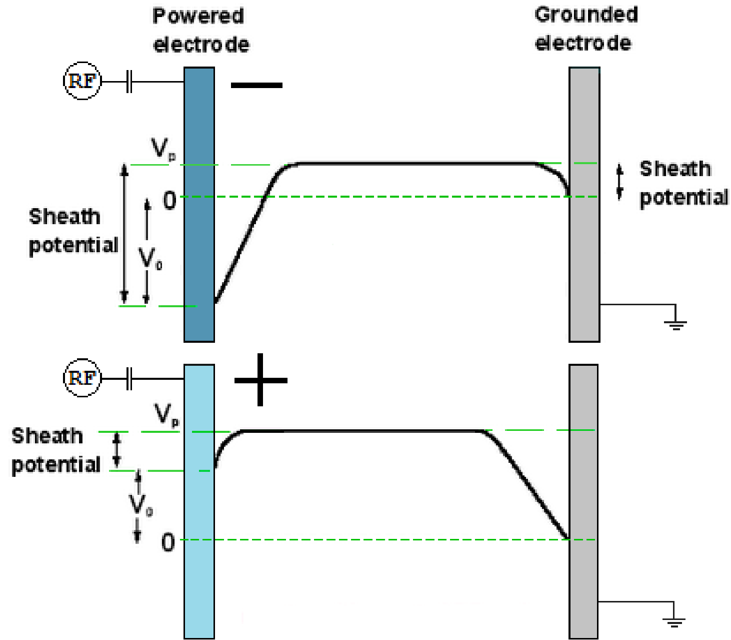


Figure 10: Plasma potential at the most negative part (top) and at the most positive part (bottom) of the RF cycle in a capacitively coupled plasma reactor with equal sized electrodes (after P. May [11]). Note that the plasma bulk is always the most positive part of the system. At the interface - plasma bulk to sheaths - the potential suddenly starts to drop the closer one comes to the electrodes.

Actually, not only the electrodes produce sheaths, but also any solid isolated object like the substrate, because much more electrons are lost to it in comparison to the heavy ions. While the potential difference between such objects and the plasma is typically about 10-25 V, it is much bigger at the electrodes, where it can easily reach several hundred volts [1].

3.2.3 Impact of the excitation frequency to the plasma properties

Naturally the excitation frequency has a strong impact on the plasma properties. However, this strong impact requires a strong change of the excitation frequency. This mainly means that small to medium variations will not be noticed so much.

Assuming typical voltages of a few hundred volts and a typical electrode distance of a few centimeters the electrons are sucked within 20 to 100 μs, which corresponds to a frequency of 10000 to $50000\,\mathrm{Hz}$. Thus, an excitation frequency of $f << 10\,\mathrm{kHz}$ may be treated as DC. In this regime the plasma density is subject of time dependent fluctuations. At very low frequencies it just turns on and off each cycle. On the other hand at frequencies of $f >> 100\,\mathrm{kHz}$ the electrons can not be sucked within one half cycle and the plasma density may be treated as constant. [1]

Note that the electron plasma frequency [1] is in the typical range $0,1 - 1\,\mathrm{Ghz}$, assuming a electron density of $10^8 - 10^{10}\,\mathrm{cm}^{-3}$. Since the excitation frequency is generally lower the electrons always will directly follow the field oscillations if the excita-

[1] When electrons are displaced against the heavy ions in a plasma there appears a Coulomb force that pulls the electrons back and make them oscillate around zero position. The corresponding frequency is called the plasma frequency $\omega_p = \sqrt{n_e e^2/(m \times \epsilon)}$.

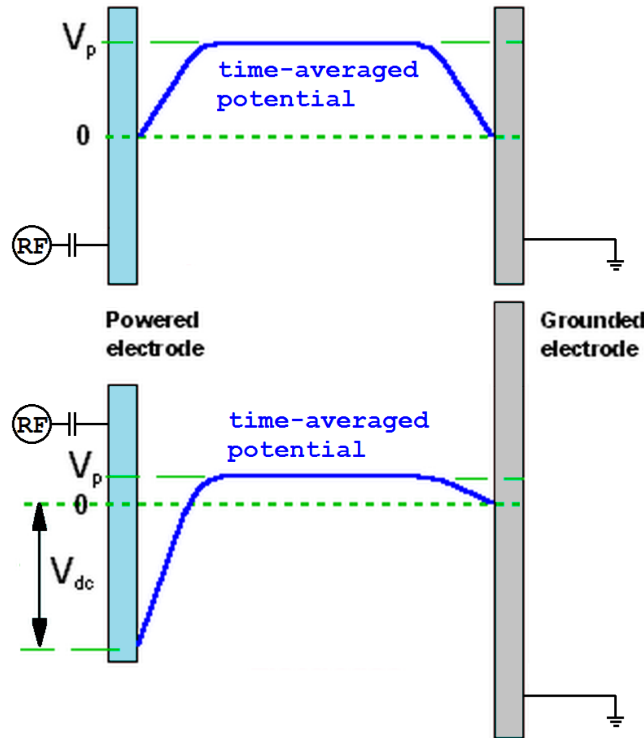


Figure 11: Time-averaged plasma potential at a capacitively coupled plasma reactor with non-equal-sized electrodes at the top and non-equal electrodes at the bottom (after P. May [11]). At the top figure the drop of the time-averaged plasma potential is equal for both sheaths, since the areas of the electrodes are the same. At the bottom figure the driven electrode is smaller. Consequently, the potential drop is larger there, corresponding to (66).

tion frequency is roughly under 1 GHz [12]. At least this will be valid at the common 13.56 MHz.

For ions one obtains other conditions. Since a sheath region has a typical thickness of $0,5-2,5\,\mathrm{mm}$ the time necessary for passing it is about $0,5-2,5\,\mu\mathrm{s}$, corresponding to a frequency of $0,4-2\,\mathrm{MHz}$. If the excitation frequency is in this regime or below, the ions can cross the sheath region within one cycle, where their energies reflect the instantaneous potential. Thus, their energy spectrum can contain virtually all energies from zero to the energy gained by the maximum potential as depicted in figure 12. On the other hand at frequencies above this region, like the very common $13,56\,\mathrm{MHz}$, the ions need several cycles to cross the sheath region, so that the energy spectrum is determined rather by the time-averaged potential. This leads to a much smaller energy band as depicted in figure 12. [1]

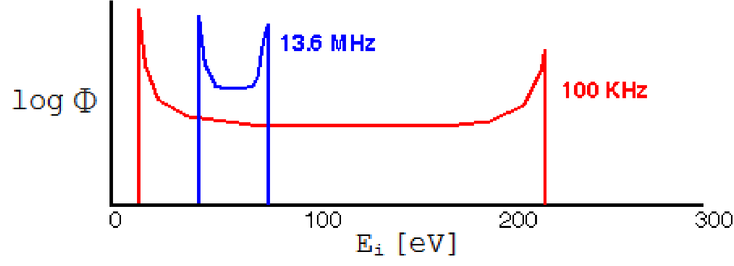


Figure 12: Calculated ion energy distributions for both excitation frequencies $100\,\mathrm{KHz}$ and $13,6\,\mathrm{MHz}$ (after [6], data from Roosmalen et. al [13]). While at low excitation frequencies the energy band of the ions is extremely broad, it gets much narrower at higher excitation frequencies.

We obtain the same finding by considering the corresponding ion plasma frequency, typically between $0,3 - 3\,\mathrm{MHz}$, again assuming an ion density of $10^8 - 10^{10}\,\mathrm{cm}^{-3}$. At this frequency the ions instantaneously follow the electric field oscillations while above these frequencies it will be governed exclusively by the time-averaged potential.

Another mechanism that is influenced significantly by the excitation frequency is the way of coupling the input power to the electrons, which is actually necessary for keeping the plasma running. At frequencies below $10\,\mathrm{kHz}$ the mechanism is similar to the one in DC plasmas. Secondary electrons are accelerated so much in one direction that they can ionize atoms or molecules. However, quite large sheath voltages of about $400 - 700\,\mathrm{V}$ are necessary to enable this mechanism. Additionally, a relatively large part of the electron energy is lost to non-ionizing collisions. [1]

At excitation frequencies in the MHz regime, two new energy transfer mechanism arise. On the one hand the high frequency AC voltage leads to displacement currents, which can be calculated equivalently to an AC current through a capacitor ($I = U\omega C$). Since the power is $P = UI$ and the current I is proportional to the frequency ω, the power dissipated due to this mechanism is proportional to the square of the frequency. In fact, this energy transfer mechanism becomes rather dominant at high pressures and large electrode distances. [1]

The second mechanism, which becomes relevant at the MHz regime, is caused by the variation of the sheath thickness within one RF cycle. A sheath of a typical size of a few millimeters moves at a speed of $10^3 - 10^4\,\mathrm{m/s}$, assuming the typical excitation frequency of about $13,56\,\mathrm{MHz}$. That corresponds pretty close to the thermal velocity of the electrons. Consequently the electrons gain energy by scattering from the sheaths. This happens because the number of electrons running into the sheath is lower if the sheath is moving away from them than if the sheath is moving towards them. So more electrons gain energy than loose energy by scattering from the sheath. [1]

Difference between floating potential and the general time-averaged sheath potential

The potential drop necessary for keeping the electron and ion flux in balance at the sheaths is called floating potential given by Chapman [14] as

$$U_f = \frac{\mathrm{k_B} T_e}{2e} \ln\left(\frac{m_i}{2,3 m_e}\right), \tag{67}$$

where k_B, T_e and e are the Boltzmann constant, the electron temperature and the elementary charge, respectively, while m_i and m_e are the ion and electron mass.

However, in case of an RF excitation, the sheath potential is RF modulated and with it the electron current, hence the electrons follow the field oscillations instantaneously (assuming $\omega << \omega_{p,e}$). If now the excitation frequency ω is $\omega_{p,i} < \omega$ the ions cannot follow the field oscillations so that they need several RF periods to pass the sheath regions. As already mentioned they follow the time averaged potential in that case. Since under that conditions the time-averaged electron current would exceed the total ion current due to the non-linear I-V characteristic [12], an additional DC offset is acquired that keeps the system in balance. The floating potential plus this additional DC offset gives the time-averaged sheath potential, eventually. So overall, in an RF coupled capacitively plasma where $\omega_{p,i} < \omega << \omega_{p,e}$, the averaged sheath potential is always higher than the floating potential which would occur in the quasi DC driven plasma. [12]

3.2.4 Sheath model

In order to model the sheaths in a capacitively driven plasma an equivalent electrical circuit has to be assumed [15]. While the plasma bulk is represented by a resistor, every sheath must be represented by a capacitor, a resistor and a diode parallel to each other (figure 13). The diodes provide that the plasma bulk has always a positive potential in comparison to the electrodes [12]. Now the two different cases of an AC plasma with a high excitation frequency is treated. The following derivations and explanations are mainly discussed following the work of Snijkers [12].

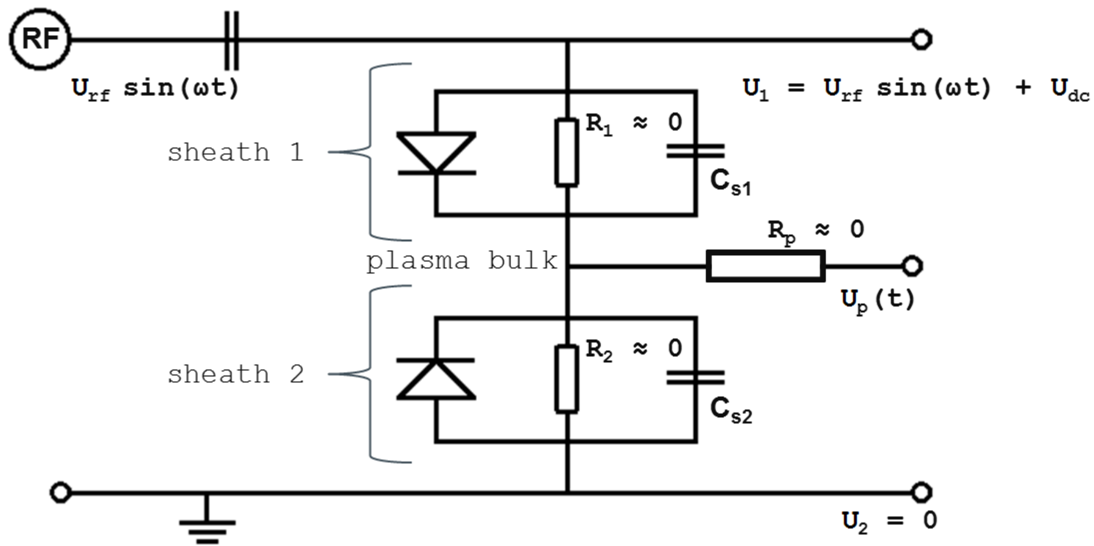


Figure 13: Equivalent circuit of an capacitively coupled RF plasma (after [15]). Since a high excitation frequency is considered, the contribution of the three resistors can be neglected. The additional DC voltage behind the capacitor results from the different sizes of the driven and the grounded electrode.

Generally it can be assumed that at high excitation frequencies the sheaths behave in a very capacitive way [16], so that the resistors at the sheaths in the equivalent electrical circuit can be canceled out [17].

As depicted in figure 13 the effectively applied voltage and the RF driven electrode (1) consists of two parts. While the first one is trivially the RF part with ΔU_{rf} and ω as the amplitude and the excitation angular frequency, the second one is an DC voltage U_{dc}, which results from the auto-bias voltage caused by differently sized electrodes,

$$U_1(t) = \Delta U_{rf} \sin \omega t + U_{dc}. \tag{68}$$

For the other part the voltage of the plasma bulk $U_p(t)$ can be described by the time-averaged potential $\bar{U}_p$ and addition modulations of the amplitude ΔU_p,

$$U_p(t) = \bar{U}_p + \Delta U_p \sin \omega t. \tag{69}$$

Since the charges in a series connection of capacitors are always of the same size, a relationship between both amplitudes U_{rf} and ΔU_p can be derived,

$$Q_2 = Q_{ges} \Rightarrow C_2 \, \Delta U_p = \left(\frac{1}{C_1} + \frac{1}{C_2} \right) \Delta U_{rf} \Longrightarrow \Delta U_p = \frac{C_1}{C_1 + C_2} \Delta U_{rf}. \tag{70}$$

As already discussed the time-resolved plasma potential is always more positive than that of the electrodes. Concerning the driven electrode, this corresponds to

$$\bar{U}_p + \Delta U_p \sin \omega t \geq U_{dc} + \Delta U_{rf} \sin \omega t, \tag{71}$$

while - with respect to the grounded electrode - the plasma potential has always to be larger than or equal to zero,

$$\bar{U}_p - \Delta U_p \sin \omega t \geq 0. \tag{72}$$

Once per period the plasma potential has to be the same as the potential of both electrodes in order to balance the net current through the sheath. Thus, both equations above can be expressed as (because of $\Delta U_p < \Delta U_{rf}$)

$$\bar{U}_p + \Delta U_p = U_{dc} + \Delta U_{rf}, \tag{73}$$

$$\bar{U}_p - \Delta U_p = 0. \tag{74}$$

This results in

$$\bar{U}_p = \Delta U_p = \frac{C_1}{C_1 + C_2} \Delta U_{rf} \tag{75}$$

as well as

$$2 \frac{C_1}{C_1 + C_2} \Delta U_{rf} = U_{dc} + \Delta U_{rf} \tag{76}$$

$$\Rightarrow U_{dc} = \frac{C_1 - C_2}{C_1 + C_2} \Delta U_{rf}. \tag{77}$$

U_{dc}, which corresponds to the auto-bias, is negative if $C_2 > C_1$. In that case the capacity of the driven electrode (1) is smaller.

Now the potential drops U_1 and U_2 of both sheaths can be calculated with respect to the amplitude of the applied RF voltage U_{rf} with help of the equations (68), (69), (77), (75) and (77),

$$\begin{aligned} U_2(t) &= U_p(t) && (78)\\ &= \bar{U}_p + \Delta U_p \sin\omega t = \frac{C_1}{C_1 + C_2}\Delta U_{rf} + \frac{C_1}{C_1 + C_2}\Delta U_{rf} \sin\omega t && (79) \end{aligned}$$

$$\begin{aligned} U_1(t) &= U_2(t) - U_{\text{driven}}(t) && (80)\\ &= U_2(t) - \left(\frac{C_1 - C_2}{C_1 + C_2}\Delta U_{rf} + \Delta U_{rf} \sin\omega t\right) && (81)\\ &= \frac{C_2}{C_1 + C_2}\Delta U_{rf} - \frac{C_2}{C_1 + C_2}\Delta U_{rf} \sin\omega t. && (82) \end{aligned}$$

Note that the voltage drop at the smaller electrode (smaller capacity) is the largest, which results in an enhanced ion bombardment. A situation which is exploited in practical applications, since the substrate is placed preferably in front of a smaller electrode. This does not only reduce the applied RF voltage ΔU_{rf}, but also conserves the material of the other electrode, which is also exposed to ion bombardment.

Although the time-averaged sheath capacities C_1 and C_2 can be calculated mainly by the effective area of the electrodes, they are also influenced by the corresponding voltage across the sheaths to some extent [15]. Obviously there exists a connection between the auto-bias voltage and the amplitude of the excitation voltage, which is treated in several theoretical works listed in [12] showing good but not perfect agreement with respect to practical experiments.

As the sheath model has been treated and simplified for the capacitive case in the high excitation frequency regime, where the resistors could be neglected, it also can be used for a quasi DC plasma, which corresponds to a purely resistive behavior [18]. It could be shown that in this case the RF sheath potential behaves non-sinusoidal [19,20]. A mathematical summary of corresponding relationships is provided in [12]. Hence this frequency regime is irrelevant for HDP CVD applications, it is no longer treated in this work.

3.2.5 Sheath ion flux

In order to quantify the steady state flux Φ of ions bombarding a surface under a certain time-averaged voltage U [1], it has to be considered firstly that the flux has to be constant everywhere in the sheath,

$$n_i(x)\, v_i(x) = n_{i,0}\, v_{i,0}, \tag{83}$$

where $n_{i,0}$ and $v_{i,0}$ are the ion densities at the edge of the plasma bulk (compare figure 14). The ion velocity $v(x)$ can be expressed by the kinetic energy, gained by dropping the distance x through the potential U,

$$E_{kin} = Q\,U = -\frac{1}{2}\, m_i\, (v_i(x)^2 - v_{i,0}^2), \tag{84}$$

where Q is the ion charge and m_i its mass. Combining these two equations results in

$$n_i(x) = n_{i,0} \sqrt{\frac{1}{1 - \frac{2QU}{m_i v_{i,0}^2}}}. \tag{85}$$

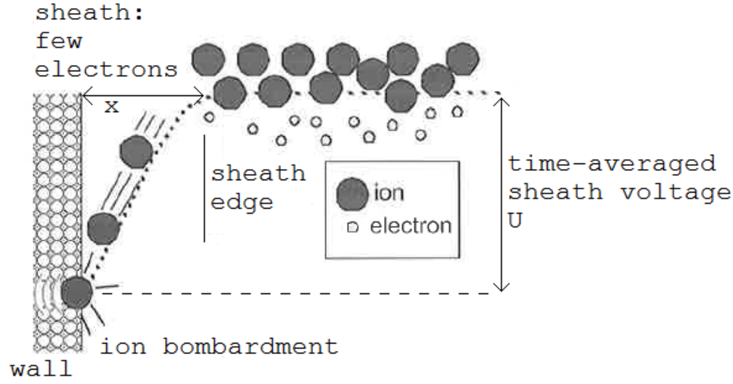


Figure 14: Schematic view of the sheath region (after Zuraw and Dobkin [1]).

In order to determine the actual potential, the electron density in the sheath region is used, which can be obtained by assuming that it is governed by Boltzmann statistics due to the electron's thermal movement,

$$n_e(x) = n_{e,0}\, e^{-\frac{E_{kin}}{\mathrm{k_B} T_e}} = n_{e,0}\, e^{\frac{qU}{\mathrm{k_B} T_e}}, \tag{86}$$

where q is the elementary charge, k_B the Boltzmann constant and T_e the electron temperature. It can be assumed that in the sheath region more ions than electrons exist in general, because otherwise a local negative charge would establish that would press the electrons towards the strongly positive plasma bulk. So the condition is given that the electron density decreases faster than the ion density towards the sheaths to the electrodes,

$$\frac{\partial n_e(x)}{\partial x} > \frac{\partial n_i(x)}{\partial x}. \tag{87}$$

This leads to

$$n_{e,0}\, e^{\frac{qU}{k_B T_e}} \frac{q}{k_B T_e} \frac{\partial U}{\partial x} > n_{i,0} \sqrt{\frac{1}{1 - d\frac{2QU}{m_i v_{i,0}^2}}} \frac{Q}{m_i v_{i,0}^2} \frac{\partial U}{\partial x}. \tag{88}$$

By considering equation (88) at $x = 0$, where the potential U can be set as zero, one finds the simple form

$$\frac{n_{e,0}}{\mathrm{k_B} T_e} > \frac{n_{i,0}}{m_i\, v_{i,0}^2}. \tag{89}$$

Additionally, it has been assumed that only single ionization is given, so that the ion's charge Q is simply the elementary charge q, with exception of the sign. In fact this is a valid assumption on the basis of the common parameters for PECVD processes.

Now the condition for the ion velocity $v_{i,0}$ - to satisfy positive charge throughout the sheath - can be determined,

$$v_{i,0} > \sqrt{\mathrm{k_B} T_e}\, m_i =: v_{\mathrm{Bohm}}. \tag{90}$$

The minimal velocity, which is necessary to fulfil this condition is called Bohm velocity v_{Bohm}. Using it the ion flux Φ_i as well as the ion current J_i to the surface one can write

$$\Phi_i = n_{i,0}\, v_{\mathrm{Bohm}}, \quad J_i = q\, n_{i,0}\, v_{\mathrm{Bohm}}. \tag{91}$$

Note that a non-zero Bohm velocity presuppose a kind of pre-sheath region, where the ions get accelerated to this velocity. The potential difference has to be on the order of $k_B T_e$ considering the Bohm velocity in (90). [1]

3.2.6 Plasma ignition: Paschen's law

Since in every non-perfect vacuum there are always some free ions and electrons, it is naturally possible to ignite a plasma with just the two electrodes, which are already there for driving a capacitively driven plasma. However, the break-through voltage, which is necessary for igniting the plasma, is strongly dependent of the product of the chamber pressure and the distance between the two electrodes.

The law which describes the relation between break-through voltage V_b, pressure p and the electrode distance d is known as the Paschen's law

$$V_b = \frac{Apd}{\ln pd + B}, \tag{92}$$

where p and d are the chamber pressure and the distance between the electrodes. A and B are coefficients that depend on the gas composition. The Paschen curve is depicted in figure 15 for some common gases.

Note that the minimum of the breakdown voltage is about $1\,\mathrm{Torr} \cdot \mathrm{cm}$. Towards smaller values the electrons cross the chamber more and more without collisions. In that case more electrons are lost than produced, so that a higher voltage is required for ignition. At higher products of pressure and electrode distance, the collisions get increasingly non-ionizing, since the mean free path decreases and the electrons do not gain sufficient energy required for ionization.

3.3 Inductively driven plasmas

3.3.1 Construction and principle

In order to drive a plasma inductively, a time-varying electrical current is provided through a coil outside the chamber so that a time-varying magnetic field is generated within the plasma that induces an electrical current there. The principle of energy transfer is similar to a transformer. The electrical current in the plasma inevitably leads to an electron heating which leads to ionization, eventually. The way the coils are wrapped around the reactor depends mainly on its form, since the only purpose is that the magnetic field penetrates the reactor. For a cylindrical reactor type the coil is usually wrapped solenoidally around the chamber. A schematic view of an inductively coupled plasma reactor is provided in figure 16.

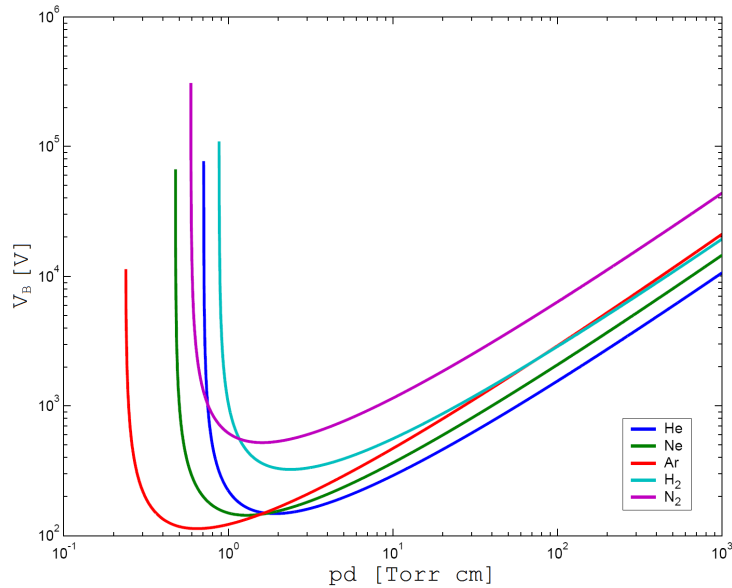


Figure 15: Breakdown-voltage as function of the product of pressure and the distance between the two electrodes (after [21]). This function is called the Paschen curve and is depicted for helium, neon, argon, hydrogen and nitrogen. The data for plotting these curves are taken from Liebermann and Lichtenberg [22].

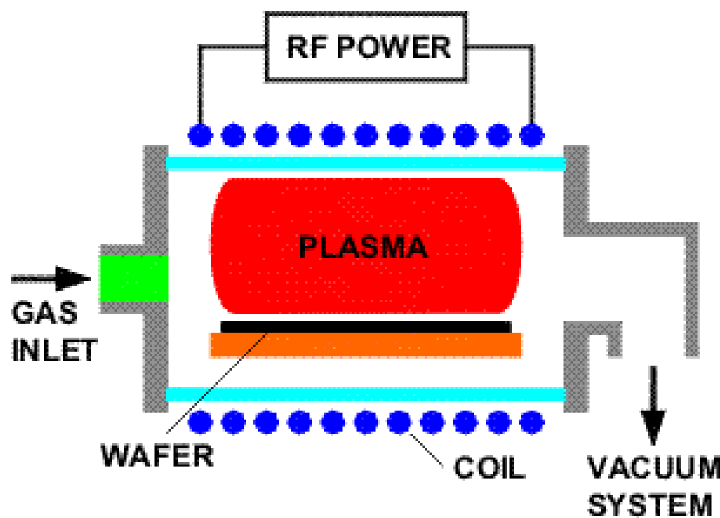


Figure 16: Schematic view of an inductively coupled RF plasma reactor (after [10]).

Since the induced currents are circular currents, no potential drops are caused by them within the chamber. Thus, the induced power can be increased, so that much more power can be coupled into the plasma without facing any voltage limitations like in a capacitively coupled plasma [1]. So overall, inductively coupled plasma reactors are usually designed as plasma tools with high plasma density ($10^{10} - 10^{12}\,\mathrm{cm}^{-3}$) at about pressures lower than $< 10\,\mathrm{mTorr}$ for deposition as well as etching purposes [23]. Nevertheless, usually capacitively coupled RF power in form of two electrodes (mostly called RF bias) is added for providing additional ion bombardement. Since the inductively part is then mainly responsible for the plasma density while the capacitively part is mainly responsible for the ion bombardment, both properties can be well adjusted.

3.3.2 Magnetic skin effect

Similar to the exponential decay of an high frequency AC current in a metal due to some skin depth, there is also an exponential decay of the magnetic field in an inductively driven plasma. Assuming the excitation frequency ω is much lower than the collision frequency ν [24], the magnetic skin depth δ depends on the excitation frequency ω, the permeability μ_0 and the plasma conductivity σ, which itself depends on the plasma density as well as the chamber pressure [1],

$$\delta = \sqrt{\frac{2}{\omega \mu_0 \sigma}}. \tag{93}$$

This corresponds to the classical skin effect. The skin depth is typically in the centimeter range in this application and can be increased by decreasing the excitation frequency that is typically at a few MHz. Alternatively, it can be increased by decreasing the chamber pressure, since this would lower the conductivity. The magnetic skin effect influences not only the distribution of the electromagnetic field, but also the way of electron heating as well as the power absorption of the plasma [24].

If the electrons move a distance comparable to the skin depth during one field period without colliding other particles, the conductivity within the skin layer becomes a function of only the RF field [24]. This case is referred to as anomalous skin effect. The amplitude of the magnetic field does not show the simple exponential shape any more. It is more complicated and not monotonic [25]. In addition, the energy dissipation of the RF energy within the plasma does not depend on the collision frequency then ν [24]. It can even exist at $\nu = 0$. This mechanism of collisionless heating works because of resonant interaction of the magnetic field with the plasma electrons and is described further by Kobolov [26] or Tyshetskiy [25], etc. Many typical ICP reactors are operating in a near-collisionsless regime where the anomalous skin effect is given.

3.3.3 Uniformity

Since the uniformity of the deposition or etch rate - depending on the inductively driven reactor's purpose - is correlated with the plasma uniformity, it is of interest how the radial density of the plasma behaves under certain conditions in an inductively coupled plasma. Especially the question to what extent the arrangement of the external coils influences the plasma is of great interest.

It could be shown by a two-dimensional fluid model of Stewart et. al [27] that the uniformity increases with a decreasing pressure in general. This is meaningful since the mobility of electrons and ions is increased and so a better energy distribution can be provided. Especially at a pressure of $p \leq 5\,\mathrm{mTorr}$ the position of the power coupling parts have a very low influence on the plasma density profile.

It could also be proved - assuming a cylindrical geometry - that the aspect ratio R/L (R radius, L height of reactor) has also a strong influence on the radial plasma density. If R/L is decreasing the uniformity of the radial plasma density is improving.

Two-dimensional models concerning the uniformity of various plasma parameters are provided for example by Economou et. al [28], DiPeso et. al [29], Vent et. al [30], Kushner and Vasenkov [31]. A three-dimensional model which considers asymmetries in a reactor (caused by pumping, etc.) is provided i.e. by Kushner [23].

3.3.4 Plasma ignition

The ignition of an inductively driven plasma suffers from the fact that almost no free electrons are present. Assuming a well adjusted matching network, the power inductively stored in the coils becomes very large, generating a large voltage across the coil, which is partly capacitively coupled to the inner wall of the chamber, eventually. The generated electric field causes a breakdown, when it is large enough. The so ignited plasma operates in a quasi-capacitive mode. As the plasma density reaches a critical level where circular currents are supported the plasma switches to an inductive operating mode. This jump can be usually observed optically as a sudden increase in brightness [1].

However, if a Faraday shield is used to minimize the otherwise rather undesirable capacitive coupling by the coils, which also exists to some extent even in the inductive mode, a further independent ignition source has to be provided [1].

3.4 Microwave driven plasmas

In some applications a higher chamber pressure is advantageous, e.g. during the chamber clean where a high fluorine flux is required for providing reasonable etch rates. This can only be enabled if the pressure is at least at a few Torr. However, in these pressure regions inductive sources tend to localize the plasma near to the inductive coil and so to the wall. Consequently, the wall is locally heated very much, which leads to mechanical stress that may even cause a cracking of the chamber [32].

For this reason it makes sense to excite the plasma by electromagnetic waves. It must be considered that electromagnetic waves are reflected out of the plasma if their frequency is below the plasma frequency ω_p. Consequently, the condition for exciting a plasma by electromagnetic waves with the frequency ω_{mw} is

$$\omega_{mw} > \omega_p. \tag{94}$$

In the common case of electron densities of about $10^{10}\,\mathrm{cm}^{-3}$, corresponding to a plasma frequency of approximately $1\,\mathrm{GHz}$, the frequency of the electromagnetic wave should be sufficiently above this value. Thus, typically microwave radiation is used.

Alternatively also a sufficiently well-sized capacitive source would be applicable in this pressure region. But it shows the disadvantage that it would produce a high voltage causing local sputtering, which is not desired in some cases. In contrast a microwave plasma provides an electrodeless system without potential drops and even without metal parts inside the reactor. The microwave source can be implemented compactly even for larger reactor chambers. Nevertheless, it permits plasma densities up to $10^{11}\,\mathrm{cm}^{-3}$. On the other hand the major disadvantage of microwave driven plasmas are their non-uniformity. The typical chamber size is large compared to the typical microwave wavelength of a few GHz. Consequently, a complex mode pattern appears producing a non-uniform plasma. Thus, the precursors are often decomposed by microwave radiation before they are injected in the chamber. Overall, microwave excitation is well suited for cleaning applications because of the higher pressure and gas flow where capacitive and inductive excitation fails due to several reasons. [1]

3.4.1 Excitation mechanism and efficiency

For the energy transfer mainly the electrons are responsible, which are accelerated periodically by the electric field of the microwave. Whereas the electrons are accelerated in the first half of the wave cycle, they are decelerated in the second half. Without collisions within this cycle the energy transfer would be zero. Only if there occurs a collision a net energy transfer happens. The energy transfer is maximized if the frequency of the microwave equals the electron collision rate, which is inversely related to the chamber pressure. At a given microwave frequency fulfilling this condition (94), the chamber pressure is chosen so that the energy transfer is maximized, which is typically some Torr. [1]

3.5 Plasma damage

A potential problem of plasma processes is the plasma damage that may occur in the device as well as the chamber walls. Especially in semiconductor manufacturing, integrated circuits that are directly exposed to the plasma can be damaged very easily resulting in a poor performance and reliability. In fact, there is not only one kind of plasma damage that may occur.

3.5.1 Plasma charging due to a lateral non-uniformity in the DC plasma potential

This kind of plasma damage can occur due to the impinging ions on the wafers surface in a capacitively driven plasma if a lateral non-uniformity in the DC plasma potential above the wafer is present. During a normal RF-cycle also a large electron current can flow when the sheath voltage decreases, because the plasma sheath is periodically growing and shrinking. This is usually no problem, because the sheath potential adjusts itself so that it enables the electron current in the integrated circuit to compensate the impinging ion current, leading to non-charged surface. On the other hand, in case of a non-uniformity in the DC plasma potential a lateral electron current through the wafer layer is produced, which can cause damage by traveling through the oxide. This is critical because of the thin oxide layer where the electrons can tunnel through. If the electron energy is too high, bonds may break and subsequently degrade the function of the IC. This negative effect can be amplified by inconvenient IC structures such as a transistor that is connected to a large area of an exposed metal. The much larger amount of collected ions due to the large area of the metal causes a larger electron current through the one transistor than it would appear if the transistor would be exposed without an additional surface. [1]

3.5.2 Damage from impinging particles and photons

Especially the gate oxide is sensitive to impinging particles or photons with high energy [1]. While high energetic ions (note that normally no electrons are impinging in such a plasma) mostly cause mechanical damage in the device structures, photons can create broken bonds if their energy is high enough [33]. However, only the near surface area is exposed so that a thin film of deposited material already protects the sensitive structures.

3.5.3 Arcing

Arcing occurs when a transient current flows between the chamber wall and/or the electrodes and the surface of the wafer [34]. It also can arise between other parts of the chamber, mostly because of peak discharges caused by spikes produced by malfunction of the reactor. Some detailed examples of different kind of arcing are treated by E. Yan [35]. Concerning the HDP CVD reactor type, which is treated in detail in the next chapter, arcing between the electrodes and the wall is of great interest, since this is a problem which sometimes occurred before and during the measurements carried out in this thesis. Arcing becomes apparent when the wafer surface shows a lump of aluminum oxide fused to the surface. The arc knocks some material out of the ceramic dome that is consisting of aluminum oxide and may subsequently drop down to the wafer surface. Unfortunately, in practice the arcing event is only detectable after the process by visual examination. Since this is usually not done for every wafer it brings about a potential uncertainty that can be overlooked.

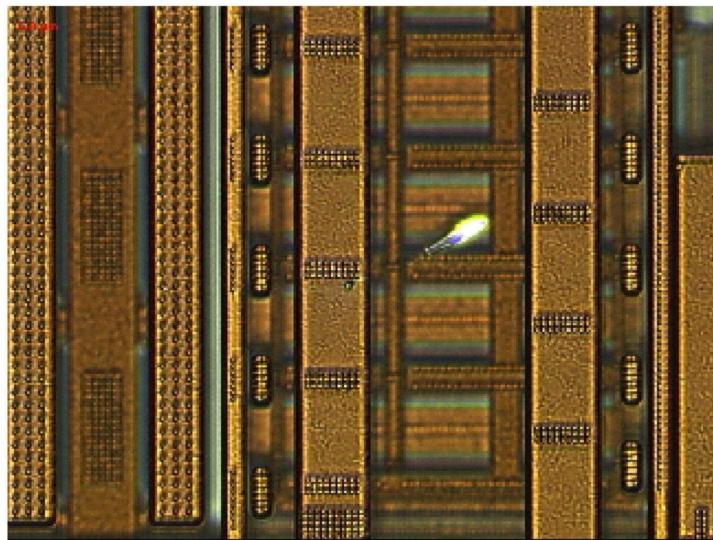


Figure 17: Typical consequence of arcs that strike the chamber wall (picture from ams AG). The aluminum containing material knocked out of the chamber wall can cause call-sign shaped layers that may damage the wafer surface (white spot in the picture).

So, regardless of the type, it would be a great advantage for process monitoring to detect an arcing event just right during the process, so that the affected wafers could be immediately sorted out and countermeasures can be performed at the machine, respectively. Actually there are two well-known approaches, which are relatively simple to implement. Optical emission spectroscopy (OES) and monitoring of the RF bias voltage that will show irregularities in case of arcing.

In practice this is not a trivial problem, because a typical arcing event takes place within a few microseconds [35]. Especially OES suffers from the fact that its is under realistic operation not possible to realize appropriate sampling rates. Furthermore, by an OES system only a limited plasma area through a small glass window can be observed, so that it is not guaranteed that every arcing event can be detected. Therefore, a high-resolution RF detector system may be the better choice, since in this case a higher sampling is easier to accomplish. Arcing events could be detected then, regardless in what area they appear.

4 High density plasma CVD

4.1 Principle

High Density Plasma Chemical Vapor Deposition (HDP CVD) is a special case of plasma enhanced CVD (PECVD). As PECVD, the HDP CVD process takes place under reduced pressure in the low pressure region, to enable a plasma state, where the electrons are much hotter than the atoms and ions, which enables a sufficiently low deposition temperature for the temperature sensitive wafers.

However, the plasma density is quite high (about $10^{11}\,\mathrm{cm}^{-3}$) and the chamber pressure is pretty low (typically around $5\,\mathrm{mTorr}$) in comparison to common shower-head reactors that are capacitively driven only. The low pressure is required to provide a mean free path that is high enough to maximize sputter rate and to minimize scattering of the ions in the sheath. Actually PECVD systems that are only capacitively driven, can not reach such high electron densities at that low pressure required for HDP CVD, since they would run into voltage limitations (compare chapter 3.3.1) and the energetic sheath electrons would simply pass right through the plasma [1], respectively. Thus, the plasma of an HDP CVD machine has to be driven inductively. That is not the only possibility to drive a HDP CVD plasma, but this will be not discussed further in this work, since the inductive excitation is very common and moreover relevant for the test machine within the framework of the thesis.

Another feature that distinguishes the HDP CVD process is the excessive ion bombardment, which actually leads to desired sputtering. This is reached by an additional capacitive excitation of the plasma. As usual, two electrode are used for that, where the driven one is directly underneath the substrate and moreover much smaller than the grounded one. This guarantees a much higher potential drop (usually called RF bias and is around several hundred of volts) according to (66) leading to a higher energetic ion bombardment at the driven electrode and thus on the substrate. On the other hand all other parts of the reactor are protected from being exposed to that higher potential drop.

The actual benefit of ion bombardment on the just deposited film caused by (partly) capacitive plasma excitation is the significantly increased density and the thereby enhanced properties of the deposited material. However, the additional excessive sputtering represents the largest difference to other PECVD processes in semiconductor manufacturing. It actually provides a void free filling of gaps with high aspect ratio (deep and thin spaces), as they occur e.g. between metal traces of semiconductor wafers. PECVD without sputtering would cause a confluence between the deposited side walls within the gaps, if the depth is much greater than the width. In other words, the edges at the top of the gaps would grow together firstly leading to voids underneath the edges. These voids would cause a severe change of the electrical properties, usually making the devices unusable.

In order to enhance the overall sputter rate without increasing RF bias, frequently an inert gas like argon is added. Under the common HPD CVD reactor conditions its ionization rate is relatively high. Additionally, the heavy ions usually have a much higher sputter cross section in comparison to the ions of precursors. So overall the sputter rate is strongly enhanced by adding argon as inert gas in the gas mixture. Moreover inert gasses are better in holding the pressure constant especially at ramping up the

process. In detail the inert gasses can temporarily replace certain missing precursors in the pre-deposition steps so that the pressure is nearly constant without changing the pumping configuration.

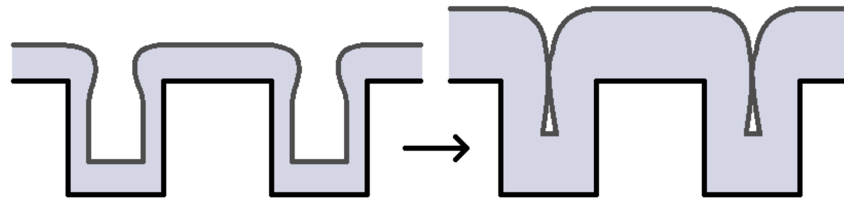


Figure 18: Schematic depiction of a SiO_2 layer deposited without simultaneous sputtering

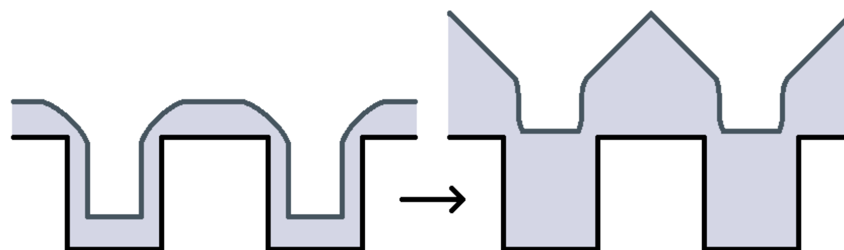


Figure 19: Schematic depiction of a SiO_2 layer deposited with simultaneous sputtering

Simultaneous sputtering makes the deposited layer to grow more quickly in the vertical direction than in the horizontal one. The reason for that is the angle dependent sputter cross section of the impacting ions (figures 20 and 21). It reaches its maximum at $45°$ while it has a minimum at $0°$ and $90°$, so that the edges are sputtered most intensively which prevents them for growing together. Due to the fact that the deposition rate excluding sputtering has its maximum at $0°$, the net deposition rate can be adjusted so that the layer grows more vertically than horizontally and even less at the edges, which finally enables the filling of high aspect ratio spaces without voids.

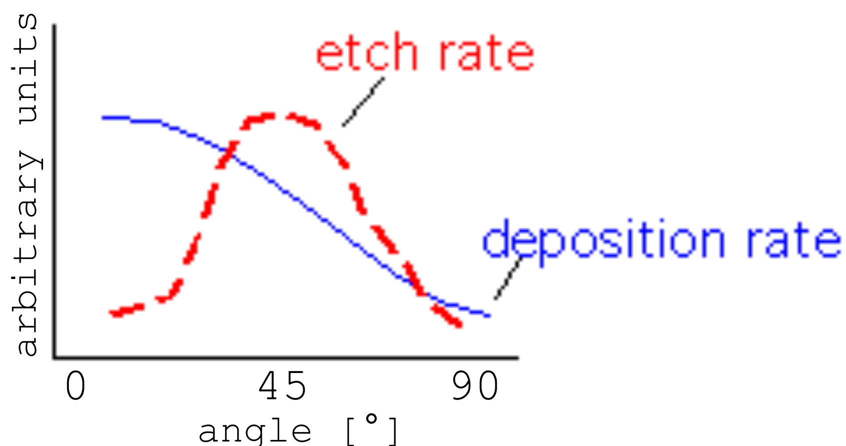


Figure 20: Angle dependent cross-section of the deposition and sputtering mechanism (after [6])

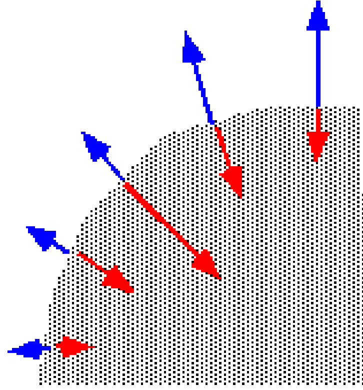


Figure 21: Schematic view of the cross-section of the deposition and sputtering mechanism (after [6])

For this reason the process is so important and inevitable in semiconductor manufacturing, where these high aspect ratios gaps are often given. However, the HDP CVD process is more expensive compared to other CVD processes, preferably because of the increased machine's complexity as well as the accompanied decreased net deposition rate. Thus, if a thick film on a wafer surface is necessary, that includes gaps with high aspect ratios, it is very common to use HDP CVD in combination with a cheaper downstream CVD process. It continues the deposition of the corresponding layer more economically as soon as the gaps are filled properly by the HPD CVD process.

4.2 Construction and operation

This chapter mainly refers to the HDP CVD reactor (Ultima HDP-CVD Centura) used within the framework of this thesis. Nonetheless the properties of this reactor are very similar to the one of other HDP CVD reactors. Details of the machine are omitted in order to maintain the confidentiality agreement with Applied Materials [2], the manufacturer of the machine. A schematic view of a typical HDP CVD reactor is depicted in figure 22.

4.2.1 Combined ICP and CCP power sources

In order to create a plasma with a sufficient plasma density, coils are wrapped around the reactor chamber to excite the plasma inductively. Here, usually a frequency around $2\,\mathrm{MHz}$ is used (so the Ultima HDP-CVD Centura). In order to provide sputtering, an electrode arrangement is used to transfer energy capacitively into the plasma with an RF frequency of typically $13,56\,\mathrm{MHz}$. Of course, this also increases the plasma density. However, it is rather a minor influence in comparison to the ICP (Inductively Coupled Plasma) source. So the main change is given in form of high energetic ions at the driven electrode, where the wafer is placed and that is smaller than the grounded one to ensure that the other parts of the reactor are not exposed to sputtering. The $13,56\,\mathrm{MHz}$ CCP excitation frequency enables a narrower band of the kinetic energy of the ions (compare chapter 3.2.3) in comparison to lower excitation frequencies. More-

[2]Applied Materials, Inc. (AMAT) is a manufacturer for products and services for semiconductor industry

over $13,56\,\mathrm{MHz}$ is reserved as industrial frequency, since these frequency regions are also used for other applications.

4.2.2 Wafer temperature and cooling

Since a large proportion of the capacitively coupled energy (around $1000-3000\,\mathrm{W}$) is given in form of high energetic ions [1], which subsequently impinge on the wafer's surface, the wafer is heated with a large proportion of this power. Unfortunately, at this low pressure only little heat would be transported away from the wafer. Consequently the wafer would quickly reach temperatures much higher than the actual gas temperature until the heat transport by radiation would stabilize the wafer temperature. However, these temperatures would be way to high for wafers with an aluminum metallization. Thus, such devices have to be actively cooled to sustain functionality. This is usually done by a backside helium cooling, where cooled helium flows around the back of the wafer. Since the helium pressure is usually about a few Torr, this results in a force that is usually compensated by an electrostatic force generated by the chuck (sometimes also mechanical clamps are used, but not so in this case).

4.2.3 Film uniformity

The quality of the deposited film on a wafer is determined by the deposition uniformity. However, a sufficient uniformity requires a constant plasma density and a constant potential across the wafer, a uniform wafer temperature and a smartly designed gas injecting system to provide an appropriate precursor distribution, etc. While a constant potential drop across the wafer can be simply reached by a suitable reactor design, an uniform wafer temperature as well as a constant plasma density above the wafer can be only provided throughout the whole life span of a reactor by designing it adjustable with respect to the crossection. The two named properties may probably change after modifications in the chamber geometry, i.e. after major maintenance interventions.

Since the wafer temperature is usually controlled by the helium backside cooling, an adjustable temperature uniformity can be provided by implementing an inner and other helium channel that can be adjusted independently. On the other hand a cross-sectional uniformity of the plasma density can be reached by using independently adjustable top and side ICP power sources as depicted in figure 22.

4.2.4 Gas flow and pumps

In order to provide high deposition rates (about $400-500\,\mathrm{nm/min}$) large gas flow rates are necessary, especially because of the low gas pressure. So for depositing a $\mathrm{SiO_2}$ layer on a $8\,\mathrm{inch}$ ($20\,\mathrm{cm}$) wafer at least $14\,\mathrm{sccm}\ \mathrm{SiH_4}$ are needed[3] [1]. However, since $\mathrm{SiO_2}$ is not exclusively deposited on the wafer, but also on all other parts of the reactor chamber, albeit with a smaller rate, the practical $\mathrm{SiH_4}$ flow is much higher and typically around $50-100\,\mathrm{sccm}$. Consequently, a combination with argon and oxygen - when $\mathrm{SiO_2}$ should

[3] sccm means standard cubic centimeter per minute. It describes the amount of gas $V(\equiv \Phi)$ at the standard pressure and temperature ($p = 1013,25\,\mathrm{hPa}$, $T = 0°\mathrm{C}$). According to the ideal gas law $pV = nRT$ the real volumetric flow $\equiv V_{flow}$ at a certain pressure and temperature can easily be calculated then.

be deposited - produces an overall flow rate of typically around $200 - 400\,\mathrm{sccm}$. This corresponds to a volumetric flow of remarkable $500 - 1000\,\mathrm{l/min}$, assuming a chamber pressure of $5\,\mathrm{mTorr}$ and a gas temperature of $1000\,\mathrm{K}$. As a consequence very powerful turbomolecular pumps are required that are very large and expensive. So actually the pressure reduction is limited by the pumps, eventually.

4.2.5 Chamber wall and cleaning procedure

Chamber cleaning is very important and necessary after every few wafers as already discussed in chapter 3. It is usually performed with the help of NF_3 that is decomposed by microwave radiation at a pressure of a few Torr (compare (3.4)). Cleaning is necessary, since otherwise the deposited layers on the chamber wall would get thicker and thicker and eventually crumble down, i.e. because of wall temperature fluctuations that produce stress to the deposited oxide layer. This would negatively affect the wafer. In order to prevent this and to reduce cleaning intervals the walls are actively heated typically slightly above $100°\mathrm{C}$. That keeps the wall temperature better constant in comparison to an unheated chamber wall and therefore reduces the danger of generation of undesired particles.

4.3 Transport analysis

Referring to the basic principle of a simple transport analysis in a CVD reactor in chapter 2, this can now be done for the HDP CVD to get an idea of its gas phase behavior. Note that it is only is a rough estimation of the test reactor parameters. However, this is sufficient in view of the simplifications that had been done concerning CVD transport analysis.

Corresponding to the test reactor the chamber shape is assumed to be cylindrical (figure 22). Furthermore, the chamber diameter as well as the height of the reactor are assumed to be $L = 35\,\mathrm{cm}$ giving an overall chamber volume of $V = 33,7\,\mathrm{l}$. As chamber pressure $p = 6,75\,\mathrm{mTorr} = 0,9\,\mathrm{Pa}$ is taken and the average gas temperature (neutrals and ions) is roughly estimated to be $T = 700\,\mathrm{K}$ based on the works of Meeks [36] and Zuraw [1]. The total gas input flow is assumed to be $\Phi_{\mathrm{in}} = 380\,\mathrm{sccm} \approx \Phi_{\mathrm{out}}$, where the proportion of SiH_4, O_2 and Ar is 25%, 40% and 35%, respectively. As net deposition rate $500\,\mathrm{nm/s}$ is taken, which is quite typical for such a system. For the deposition rate without sputtering $650\,\mathrm{nm/s}$ is a good approximation.

4.3.1 Volumetric flow

The total gas input flow Φ_{in}, the chamber pressure p and gas temperature T gives a volumetric flow of

$$V_{\mathrm{flow}} = \frac{p_s\,T}{p\,T_s}\,\Phi_{\mathrm{in}} = 2100\,\mathrm{l/s}, \tag{95}$$

where p_s and T_s are the standard pressure and standard temperature ($p = 1013,25\,\mathrm{hPa}$, $T = 0°\mathrm{C}$). Since virtually all of it ($\Phi_{\mathrm{in}} \approx \Phi_{\mathrm{out}}$) has to be removed again by the pumping system - this is actually quite a bunch - the volumetric flow (and consequently the minimum pressure at a certain total gas inlet flow) is limited by the pumping system, eventually.

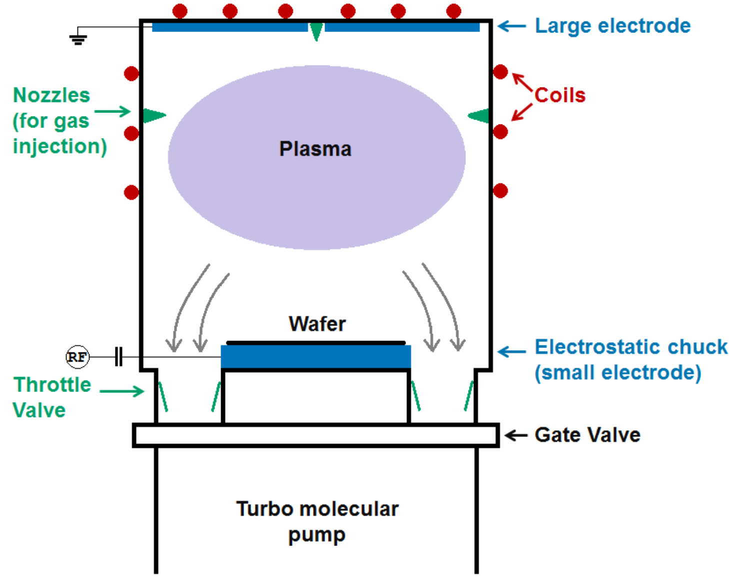


Figure 22: Schematic view of the HDP CVD reactor used for transport analysis. While the coils excite the plasma inductively and are mainly responsible for the high plasma density, the two electrodes provide a capacitive excitation and as a consequence a massive ion bombardment that causes angle-dependent sputtering. Since the electrode directly underneath the wafer is smaller than the other electrode, sputtering occurs only there because of the higher potential drop according to (66). Note that for adjusting the uniformity, the coils at the top and the side can be adjusted independently.

4.3.2 Average gas residence time and velocity

According to the volumetric flow V_{flow} (95) and the chamber volume V, the average gas residence time t_{res} can be calculated as

$$t_{\text{res}} = V/V_{\text{flow}} = 0,016\,\text{s}. \tag{96}$$

Further, the average gas velocity U can be calculated with respect to the chamber size L,

$$U = L/t_{\text{res}} = 21,9\,\text{m/s}. \tag{97}$$

4.3.3 Mass diffusivity

Assuming an average neutral particle mass of $m = 16$ and an average sphere diameter of $a = 0,5\,\text{nm}$, the mass diffusivity D is according to equation (19)

$$D \approx \sqrt{\frac{\text{k}_{\text{B}}^3}{\pi^3\,m}\,\frac{T^{3/2}}{p\,a^2}} = 46200\,\text{cm}^2/\text{s}. \tag{98}$$

Diffusion length and Peclet number
Due to equation (11) the diffusion length L_D is

$$L_D = \sqrt{4D\,t_{\text{res}}} = 61,3\,\text{cm}. \tag{99}$$

Note that this is similar to the chamber's size L, so that the Peclet number (12) is about

$$Pe = \frac{LU}{D} = 4\left(\frac{L}{L_D}\right)^2 = 1,3. \tag{100}$$

Therefore, the HDP CVD mass transport is not only diffusion dominated, but also not exclusively convection dominated.

4.3.4 Surface reaction rate constant

Assuming a deposition rate (excluding sputtering effects) of $r_{\text{depo}} = 650\,\text{nm/min}$ the surface reaction rate constant K_s can be estimated according to the surface reaction rate R (2.4),

$$R = K_s\, A_{\text{film}}\, n_{\text{pre}}, \tag{101}$$

where A_{film} is the area, that is exposed to deposition, and n_{pre} is the approximated steady state precursor gas concentration in the reactor. On the other hand the surface reaction rate R can be expressed as

$$R = n_{\text{film}}\,\frac{V_{\text{film}}}{t_{\text{film}}}, \tag{102}$$

where V_{film} is the volume of the deposited $\mathrm{SiO_2}$ film during the time t_{film}. n_{film} is the average $\mathrm{SiO_2}$ particle density within the film which is calculated by the given mass density $\rho_{\mathrm{SiO_2}} = 2.6\,\text{g/cm}^3$ of solid $\mathrm{SiO_2}$ and its molar mass $M_{\mathrm{SiO_2}} = 30\,\text{g/mol}$. n_{film} is then

$$n_{\text{film}} = \frac{\rho_{\mathrm{SiO_2}}}{M_{\mathrm{SiO_2}}}\, A_g = 5.22 * 10^{22}\,\text{cm}^{-3}, \tag{103}$$

where A_g is Avogadro's constant. If now the two equations (101) and (102) for R are set equal, one obtains

$$K_s = \frac{n_{\text{film}}}{n_{\text{pre}}}\,\frac{V_{\text{film}}}{A_{\text{film}}\,t_{\text{film}}} = \frac{n_{\text{pre}}}{n_{\text{film}}}\,\frac{d_{\text{film}}}{t_{\text{film}}} = \frac{n_{\text{pre}}}{n_{\text{film}}}\,r_{\text{depo}}. \tag{104}$$

Before calculating K_s the precursor gas concentration n_{pre} has to be determined. For that the ideal gas law can be used,

$$p_{\text{pre}}\,V = N_{\text{pre}}\,\mathrm{R}\,T_g \Longrightarrow n_{\text{pre}} = \frac{p_{\text{pre}}}{\mathrm{R}\,T_g} \approx 4\cdot 10^{13}\,\text{cm}^{-3}. \tag{105}$$

The partial pressure of the precursors (p_{pre}) was roughly estimated to be $p_{\text{pre}} = 0.43\,p = 0.39\,\text{Pa}$. This appears appropriate in view of the $\mathrm{SiH_4}$, $\mathrm{O_2}$ and Ar proportions that give a proportion of $\mathrm{SiH_4}$ and $\mathrm{O_2}$ of 65%. By taking into account the production of by-products such as $\mathrm{H_2}$ according to the overall reaction equation $\mathrm{SiH_4} + \mathrm{O_2} \rightarrow \mathrm{SiO_2} + \mathrm{H_2}$, which are subsequently always present in the gas phase, the proportion is reduced by one third, giving 43% as effective precursor proportion.

Now finally the surface reaction rate constant K_s can be calculated,

$$K_s = \frac{n_{\text{film}}}{n_{\text{pre}}}\,r_{\text{depo}} = 1400\,\text{cm/s}. \tag{106}$$

4.3.5 Average consumption time

Assuming that deposition only takes place at the wafer with the surface $A_w = \pi \cdot 10^2\,\mathrm{cm}^2$, but not on the chamber wall, the average consumption time t_{cons} is according to equation (3)

$$t_{\mathrm{cons}}^{\mathrm{wo}} = \frac{V}{K_s\, A_w} = 0.077\,\mathrm{s}. \tag{107}$$

However, since the reaction $\mathrm{SiH_4} + \mathrm{O_2} \rightarrow \mathrm{SiO_2} + \mathrm{H_2}$ is strongly exothermic (2.6), deposition also takes place at all other parts of the reactor. Thus, the exposed surface A is much larger in reality and corresponds rather to the total inner wall of the reactor, which is $A = 9620\,\mathrm{cm}^2$. This changes the consumption time dramatically to

$$t_{\mathrm{cons}} = \frac{V}{K_s\, A} = 0.0025\,\mathrm{s}. \tag{108}$$

4.3.6 Transport limited deposition rate

By comparing the average residence time $t_{\mathrm{res}} = 0.016\,\mathrm{s}$ (96) and consumption time $t_{\mathrm{cons}} = 0.0025\,\mathrm{s}$ ($t_{\mathrm{cons}}^{wo} = 0.077\,\mathrm{s}$) according to chapter (2.4) it can be stated that HDP CVD is operating in the starved regime ($t_{\mathrm{cons}} << t_{\mathrm{res}}$). The deposition rate is therefore at least partly mass transport limited.

If deposition would only take place on the wafer, the HDP CVD reactor would rather operate in the differential regime, because of $t_{\mathrm{cons}}^{\mathrm{wo}} > t_{\mathrm{res}}$. However, the deposition rate benefits rather from colder surface temperatures [36] that actually are present at all inner parts of the reactor that are not exposed to sputtering. Thus, the first approach - all inner parts of the reactor are exposed to deposition - seems the more appropriate one. Consequently the HDP CVD reactor with similar parameters as used above is assumed to be a starved reactor at normal operation.

4.3.7 Transport analysis summary

The following statements about the HDP CVD process can be summarized after the simple transport analysis:

- For the transport of the precursors both mechanisms - mass diffusion and mass convection - are of importance, since the diffusion length has a comparable size to the chamber size (Peclet number close to one). Thus, the injection of the gasses has to be performed very smartly, to ensure at least a sufficient gross distribution of the precursors by convection. As required, diffusion provides then a fine distribution, eventually.

- Since deposition of $\mathrm{SiO_2}$ takes place on all parts of the reactor, albeit with different density and quality, the deposition rate tends to be rather mass transport limited at least at the test reactor. This might make a monitoring of the concentration of certain gas phase species useful, to approximate the actual deposition rate. In fact this is the later discussed approach, where an optical emission spectrometer (OES) is used to predict the deposition rate, because the intensity of selected spectral lines are proportional to the concentrations of corresponding gas phase species.

- The requirements on the pumping system to provide such a large volumetric flow ($\approx 2000\,\mathrm{l/s}$) are quite high. In fact, this large volumetric flow is necessary to provide economically high deposition rates. Hence they tend to be mass transport limited, the feed rate of precursors as well as the removable rate of the by-product (especially H_2) is a large limiting factor. However, a higher inlet gas flow at the expense of a higher chamber pressure may be counterproductive, since the low pressure is required to provide the high sputter rate that provides the dense and conform SiO_2 deposition. The high sputter rate is strictly spoken reached by a high plasma density, which is in view of the ICP source only possible at a reasonable low chamber pressure. Moreover, a higher pressure would decrease the mean free path of the sputtering ions, preventing them from impinging perpendicular to the wafer surface, which is also necessary for conform deposition.

4.4 Process chemistry

The process chemistry of the treated HPD CVD process seems to be quite simple by inspecting the overall reaction equation $SiH_4 + O_2 \rightarrow SiO_2 + H_2$. However, by taking a closer look, this turns out to be not true. In the gas phase alone there are 100s of possible reactions species, as soon as the precursors are decomposed by high energetic electrons and naturally react with each other. Additionally, there are a bunch of surface reactions, which are actually even more complex.

A quite extensive treatment of this topic was done in the work of Meeks [36], where the majority of all significant reaction paths are listed for the gas phase as well as for the surface. Additionally the so called Aurora model [4] is applied to three different HDP CVD reactors in order to find out dominant gas phase and surface reactions paths. With the model also the molar gas composition are calculated, which can be considered as a rough estimation for the treated HDP CVD reactor within the framework of this work. Moreover, other important process parameters like the electron temperature, plasma density or the gas temperature, etc. are calculated and match the expectations.

4.4.1 Dominant gas phase reactions

The dominant gas phase reactions are depicted in figure 23 (after Meeks et al. [36]). While Argon as inert gas can be just excited or ionized, O_2 can be additionally decomposed into the radical O that can itself get excited or ionized. Further O (and sometime O_2) takes part of several Silicon-related reactions within the third path, where silane (SiH_4) is the starting point. While SiH_4 is mainly decomposed by electron collisions to SiH_3, SiH_2 and SiH, where SiH_2 and especially SiH are already very important for certain surface reaction, O and O_2 are mainly responsible for their oxidation into SiH_2O, $HSiO$ and SiO, compounds which are also needed for surface reactions.

The decomposition of the precursors is only possible because of excessive electron collisions, provided by that special plasma state where the electrons have a much higher thermal energy than neutrals and ions. In further dissociation reactions between the precursors beside O or O_2 also H or H_2 are included. Of course, there exist

[4] With the Aurora model actually a FORTRAN computer program is meant that provides a prediction of the steady-state or time-averaged properties of perfectly stirred plasma and thermal chemistry systems developed by Meeks and Moffat [37]

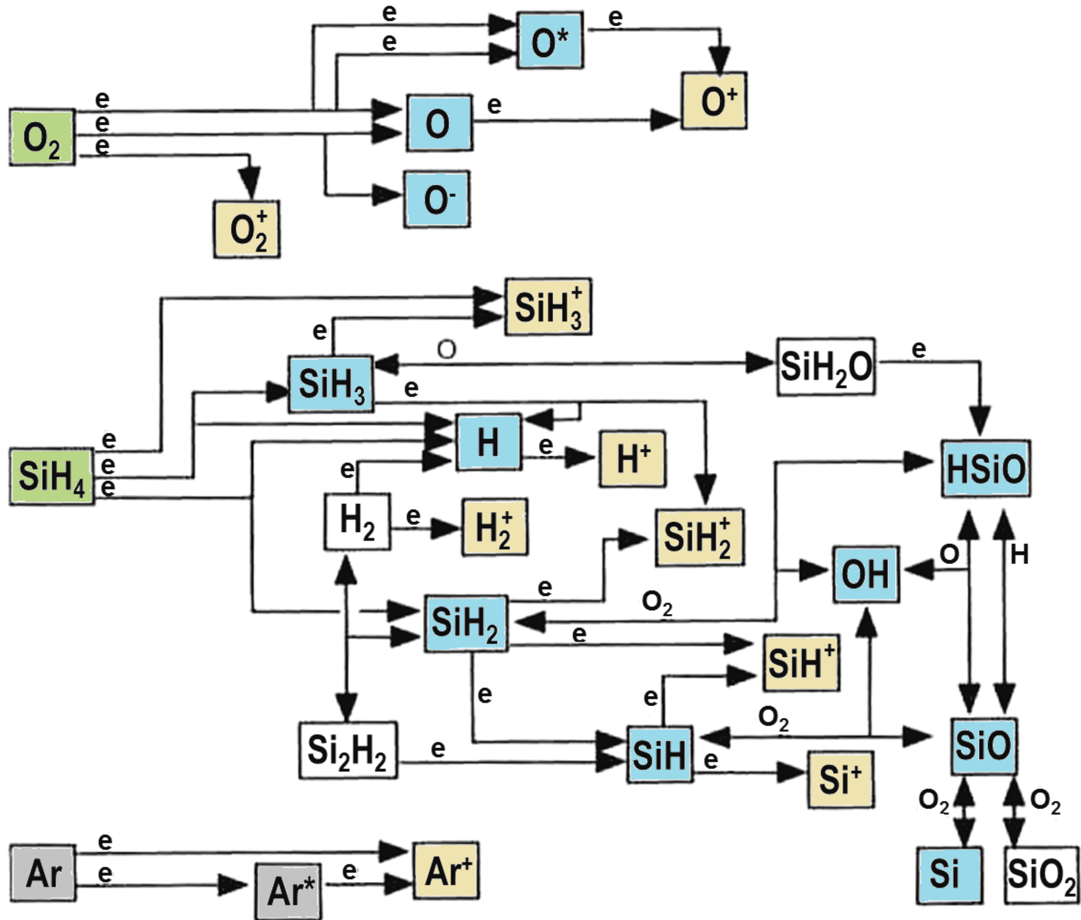


Figure 23: Schematic view of the dominant gas phase reactions of the HDP CVD reactor in the deposition mode. The green marked species indicate the precursors as they are injected into the chamber. While the decomposed precursors are marked white, the radicals and ions are marked blue and yellow, respectively.

more rather exotic reactions (listed in Meeks work [36]), but they are not that dominant like the depicted one. In almost all dissociation reactions at least one radical is included as reactant. The reason for that is discussed in chapter 2.6.2.

The estimated steady state molar composition of the gas phase is shown in figure 24. Note that the total amount of Si is much smaller than the total amount of H. It should regularly have a proportion of 1:4 according to $\mathrm{SiH_4}$, but because of the consumption of corresponding Si-compounds by the surface, the proportion is much smaller (1:11), which confirms the assumption that the deposition rate is at least partly mass transport limited 4.3.6.

4.4.2 Dominant surface reactions

The dominant surface reactions are depicted in figure 25 after Meeks et. al [36]. While the white areas indicate surface-bonded species, the grey areas mark gas phase species that participates in the surface reactions. Note that these gas phase species, that have to impinge directly from the gas phase to the surface to take part at the surface reaction, are mostly H and O.

In detail the formation of surface bounded silicon dioxide mainly leads over the sin-

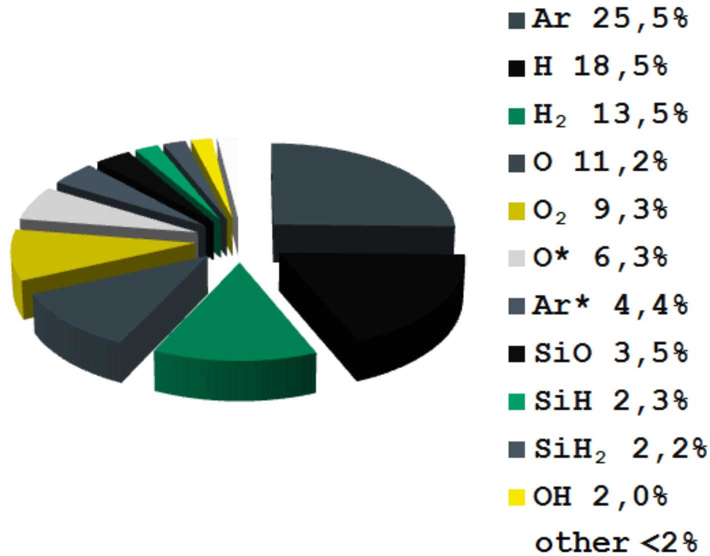


Figure 24: Approximated concentration of the most represented gas phase species of the HDP CVD reactor discussed in chapter 4.3 following the model calculations of Meeks et al. [36]. Since the proportion of injected gasses used by Meeks corresponds largely to the one of chapter 4.3 and so the power-density, it is actually a sound estimation for the HDP CVD reactor used in this work.

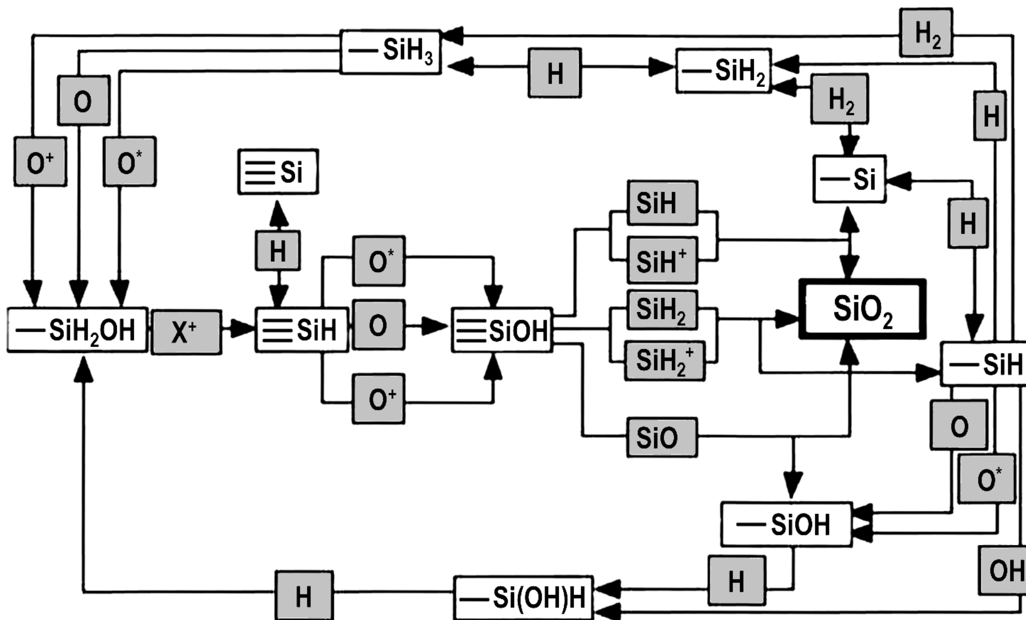


Figure 25: Schematic view of the dominant surface reactions in the HDP CVD reactor during the deposition phase after Meeks et. al [36]. While the white marked areas indicate the surface-bonded species (bonded with the underlying silicon dioxide), the grey marked areas indicate the species that participate directly out of the gas phase at the corresponding surface reaction.

gle bonded SiH_2OH that reacts with a positive ion to the triple bonded SiH, followed by an oxidation to SiOH. SiOH finally reacts by impingement of gaseous SiH_x compounds or also SiO to SiO_2. Since the gas concentration of SiH_x as well as that of SiO tends to

be low, because the deposition rate tends to be mass transport limited, the rate of this final reaction (triple bonded SiOH to $\mathrm{SiO_2}$) is likely to depend significantly on these concentrations.

It is supposed by [1], that in case of a non-correct stochiometry at the film the surface-bounded species $\equiv \mathrm{SiH}$ or $\equiv \mathrm{SiOH}$ can be incorporated in the growing film, depending on the stochiometric imbalance. In fact, at a very low oxygen to silane proportion, the film would rather consist of amorphous silicon than of silicon dioxide, because of the reaction $\equiv \mathrm{SiH} + \equiv \mathrm{SiH} \rightarrow \equiv \mathrm{Si} - \mathrm{Si} \equiv$. Generally it can be stated that the amount of incorporated hydrogen can be reduced by increasing the total ion bombardment flux, ion energy or by increasing the substrate temperature. [1]

5 Optical emission spectroscopy of HDP CVD processes

5.1 Motivation

In semiconductor manufacturing plasma processes determine the device's property and quality significantly, as many other processes. So even small deviations of the normal process behavior can cause severe deviations of the electrical properties of the wafers. Or even worse, there can occur undetected machine faults that are responsible for non-functioning chips. The later such machine faults and deviations are detected, the more money it costs, since normally not only one lot [5] is processed incorrectly then. In order to find these malfunctions at least in a reasonable time, there are regular electrical control measurements carried out after about 25% of each single process step, which is actually a significant expense. Moreover, these tests can not yet guarantee that several lots run over a machine until the fault is detected, eventually. Regarding the cost factor of far processed wafers this is even more expensive.

That is the reason why in the last years system solutions for Fault Detection and Classification (FDC) and Virtual Metrology (VM) came up. While the purpose and benefit of a FDC system is self-explanatory, this is maybe not so clear for Virtual Metrology (VM). VM enables a prediction of process results, so that small deviations of the average process result can be detected in-situ. Nonetheless, both methods provide a central data acquisition of all required parameters of the machine. That enables a rough but often sufficient fault detection and classification by simply defining lower and upper parameter limits for certain machine parameters. Maintenance actions can be triggered by these monitored parameters, leading from preventive maintenance to the more efficient predictive maintenance, where maintenance is mainly proceeded when required. However, sometimes there are more complicated models required.

VM almost always needs a more or less complicated model (mainly statistical) to provide an appropriate prediction of process results. Unfortunately, for many processes, where VM is of interest (i.a. plasma processes), the parameters collected by the machine itself are not sufficient for VM. This is the reason why in such cases external sensors are added. Their data are then collected and prepared together with the data of the machines. Depending on the type and condition of the sensor as well as of the model's quality VM might be possible, eventually. The corresponding models are mainly statistical, due to the immense complexity of the system and its many influencing parameters, where even the minor ones may have a strong influence. However, in this case a model training is required, which can be controlled then and updated by a significantly reduced number of physical measurements.

Since plasma processes are quite complex with respect to their behaviour and do not provide sufficient information for comprising FDC and VM by default (at least the test system within the framework of this thesis), it is necessary to add also additional sensors. A very simple and obvious approach is to observe the optical emission of the plasma with a sufficiently accurate spectrometer: It provides direct information about the plasma related to most machine parameters. Compared to other typical external sensors, Optical Emission Spectroscopy (OES) is an absolutely non-invasive

[5] The term lot is used in semiconductor manufacturing for a box of wafers. The machines are not fed wafer by wafer, but by these lots. This simplifies the handling of the wafer, especially because these lots inform the machine about the their content automatically.

measurement method in contrast to Langmuir probing, etc. It is mounted outside of the chamber via an optical fibre. Through a window in the chamber wall the plasma can be observed.

Since already small shifts in the chamber geometry or other process parameters may have a noticeable impact on the plasma spectrum, OES is a great tool to enhance FDC significantly. Because of the additional information about the concentration of certain gas phase species VM might be possible then. Within the framework of this thesis the prediction of the layer thickness deposited by the High Density Plasma Chemical Vapor Deposition (HDP CVD) process was of special interest because of its influence to the electrical properties of the wafers. A reliable prediction would reduce the test measurements, making the process more economically.

5.2 Experimental setup and procedure

In order to record the optical spectrum of the HDP CVD process, in a first step the spectrometer USB 4000 from Ocean Optics was used. Since its accuracy was not satisfying, the spectrometer had to be changed. A Maya2000 Pro spectrometer fulfilled the requirements (also manufactured by Ocean Optics). Thus, all spectra which are presented in this work were recorded with the Maya2000 Pro. Since the chamber had two quartz windows, the light of the chamber could be easily recorded by using an optical fiber, where one end was mounted to the sensor and the other one to the chamber window together with a collecting lens to enhance the light intensity. The arrangement was wrapped by a cover to protect it against the ingress of light from outside. The sensor was connected with a laptop for data storage. Under normal operation also an on-line-connection to the central data collection system was realized. It enabled an automatic matching between machine and sensor data and additionally the automatic application of certain models.

5.3 Optical spectrum of the HDP CVD process

Since there are various process steps until the actual deposition step with ion sputtering can start, also the corresponding optical emission spectrum varies. The most of these steps are for ramping up the process and its adjustment. The only longer one is a pure-deposition step with a duration of about 10 seconds for preparing the wafer for ion sputtering, because certain sensitive parts of the wafer surface (like the thin gates) would be damaged by the ions when they are not already covered by a sufficiently thick $\mathrm{SiO_2}$ layer. The injected gasses are always Ar, $\mathrm{SiH_4}$ and $\mathrm{O_2}$ (with exception in the cleaning procedure), however, sometimes one or two of them are missing. Anyhow the main deposition step where sputtering is included is responsible of about 90% of the deposited layer. Therefore the focus will be put on that step.

Before modeling based on OES data, the first task is to analyze the spectrum to identify as many gas phase species as possible. The corresponding gas phase species are discussed in chapter 4.4.1. Actually the observable species will be mainly atoms or single ionized atoms and also partly two atomic molecules, because at molecules with a higher number of atoms the vibrational transitions split up the observable (from $200 - 1100\,\mathrm{nm}$) electronic transitions so much that the remaining intensity of each peak will be too weak or covered by other lines. In order to monitor them, Fourier Transform

Infrared Spectroscopy (FTIR) would be necessary, hence this would enable the direct observation of the vibrational transitions in the infrared region.

The gas phase species that are observable with the OES sensor are discussed below. An optical spectrum during the main deposition step where sputtering is included being responsible for about 90% of the deposited oxide layer is depicted in figure 26. For a reliable analysis (based on line-intensity ratios) the background level has to be subtracted and the overlapping of spectral lines considered. Thus, the suspected best suited wavelength for the estimation of the background-level are also mentioned below, at least for the lines that are actively used for monitoring (boldfaced lines).

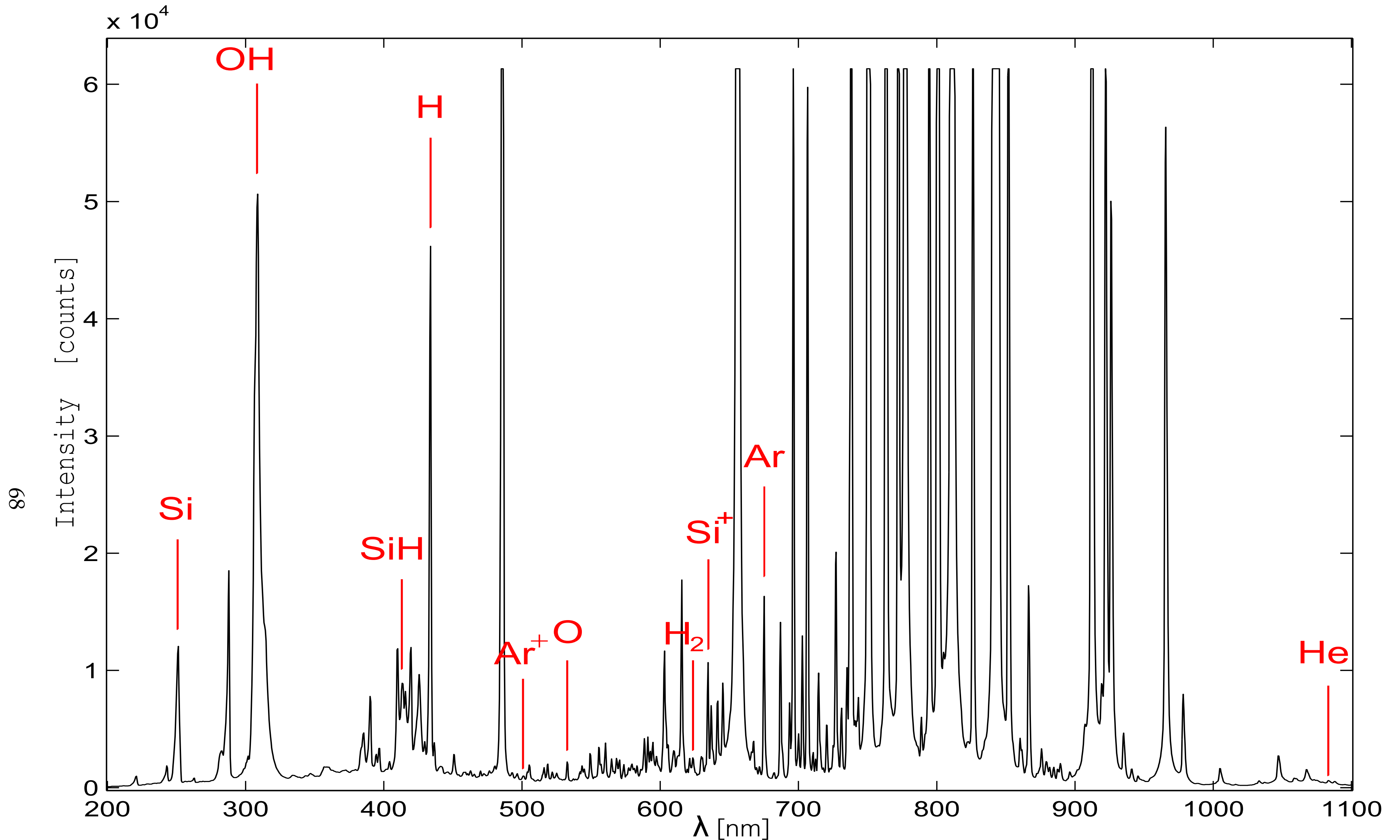


Figure 26: Optical spectrum ranging from 200 nm to 1100 nm of the HDP CVD machine plasma during the main deposition step, recorded using the Maya 2000 Pro spectrometer of ocean optics. The red marked lines correspond to the bold-highlighted wavelength within this chapter, they are selected for monitoring. They include all species found that are observable with the OES with an appropriate accuracy (compare 4.4.1).

Ar

Since neutral argon has several well observable spectral lines within the visible range and the gas mixture contains up to 40% of it, this species is comparatively simple to monitor. A few lines that can be observed quite well are listed in the following table: The Ar spectral line $675,3\,\mathrm{nm}$ is probably the best suited for monitoring, since the

Table 2: A selection of well observable spectral lines of neutral argon (Ar). The corresponding literature values were found in the NIST database [38]. The spectral line used for monitoring is bold-highlighted. $\lambda_{\mathrm{Ar,i}}$ is the wavelength of the corresponding Ar spectral line, 'ground' designates the wavelength at which the ground-level of the corresponding Ar spectral line is chosen.

$\lambda_{\mathrm{ArI,i}}$ [nm]	ground [nm]
419,8	
516,2	
518,8	
675,3	680,1
703,0	
1047	

signal-to-noise ratio is sufficiently high and the overlapping by other lines is minimal.

Ar^+

The single ionized argon atom can also be observed, despite its low gas concentration that comes from the low ionization rate about 1% (which is actually high in relation to other PECVD processes).

The first two lines are not well-suited for model building because they are strongly superimposed by the intensive H_β line at about 486 nm that is broadened quite strongly by Stark-effect. However, the third line at 501,4 nm does not show this problem.

H

Neutral hydrogen (H) stems from the precursor silane (SiH_4) and has a high concentration in the gas phase. A few lines of its Balmer-series are observable within the sensor range. The two most intensive lines H_α and H_β normally saturate the sensor, because usually the integration time is adjusted to monitor also weak lines of some species. The lines with an appropriate intensity in that context are $H_\gamma\,434,0\,\mathrm{nm}$ and $H_\delta\,410,1\,\mathrm{nm}$, where H_δ is partly superimposed by other lines. Hence $H_\gamma\,434,0\,\mathrm{nm}$ is used to monitor neutral hydrogen. The ground it is determined at the wavelength $431,2\,\mathrm{nm}$.

O

Oxygen (O) is also strongly contained in the gas phase, actually even more than O_2.

Table 3: A selection of well observable spectral lines of the single ionized argon atom (Ar^+). The corresponding literature values were found in the NIST database [38]. The spectral line used for monitoring is bold-highlighted. $\lambda_{ArII,i}$ is the wavelength of the corresponding Ar^+ spectral line, 'ground' designates the wavelength at which the ground-level of the corresponding Ar^+ spectral line is chosen.

$\lambda_{ArII,i}$ [nm]	ground [nm]
476,5	
480,6	
501,4	499,1

Well recognizable lines are:

Table 4: A selection of well observable spectral lines of neutral atomic oxygen (O). The corresponding literature values were found in the NIST database [38]. The spectral line used for monitoring are boldfaced. $\lambda_{O,i}$ is the wavelength of the corresponding O spectral line, 'ground' designates the wavelength position of the corresponding O spectral line, where its value is determined.

$\lambda_{O,i}$ [nm]	ground [nm]
532,9	531,0
615,67	

OI $532,9\,\mathrm{nm}$ was finally used for monitoring, since the influence of other lines turned out to be the smallest.

Si
Atomic neutral silicon (Si) coming from silane SiH_4, has a gas phase concentration lower than 1% according to Meeks et. al [36]. Nevertheless, it can be relatively easily observed in the OES-range: Since the line at $390,4\,\mathrm{nm}$ is probably significantly superimposed by a spectral line of $Si+$, the other line at $251,4\,\mathrm{nm}$ is used for monitoring. Actually this line consists of several Si lines according to the NIST database [38], which, however, is no reason for exclusion concerning monitoring.

Si^+
Single ionized silicon (Si^+ or $SiII$) is according to the ionization degree of about 1% 100 times lower concentrated than non-ionized silicon (Si). Thus, it is remarkable

Table 5: A selection of well observable spectral lines of neutral atomic silicon (Si). The corresponding literature values were found in the NIST database [38]. The spectral line used for monitoring is boldfaced. $\lambda_{\text{SiI,i}}$ is the wavelength of the corresponding Si spectral line, 'ground' designates the wavelength position of the corresponding Si spectral line, where the ground-level its value is determined.

$\lambda_{\text{SiI,i}}$ [nm]	ground [nm]
390,4	
251,4	254,7

that nonetheless an observable line could be found within the HDP CVD spectrum at $634,7\,\text{nm}$. For the ground-determination the wavelength $632,1\,\text{nm}$ was chosen.

H_2

Hydrogen (H_2) has a weak but observable line at λ_{H_2} $623,7\,\text{nm}$ [39], which is actually a line band because of the vibrational transitions that can not be resolved by the used spectrometer configuration. There is probably also an overlap with other lines. For the ground-determination the wavelength 622,8 nm was chosen.

OH

The molecule OH has a very intensive line band at λ_{OH} $308,8\,\text{nm}$ [40], and could therefore easily be monitored (ground-determination at wavelengths $292,3\,\text{nm}$ and $331,1\,\text{nm}$).

SiH

The Si-molecule SiH that plays an important role for surface reactions, has a moderately observable band at about $413,8\,\text{nm}$ [41] (ground-determination at the wavelengths $411,7\,\text{nm}$ and $417,6\,\text{nm}$). Unfortunately, the influence of other lines can not be excluded. However, it is the only possibility to monitor SiH within the $200-1100\,\text{nm}$ range.

SiO

Silicon-oxide SiO that also plays an important role for surface reactions, even though not so much as SiH, is hardly observable in the sensor range. Only the line band at $234,6\,\text{nm}$ [42] occurs there, unfortunately extremely weak at the used sensor integration time. Moreover, the signal-to-noise ratio is very bad, also because of the low sensitivity of the OES-sensor in this range. Therefore this line band is definitely not suited for an appropriate monitoring of the species.

O_2

Actually there should be an observable line band for oxygen O_2 at $760,5\,\text{nm}$ [43]. However, it is covered by a very intense argon line at $763,5\,\text{nm}$.

He
Helium is only contained in the gas phase due to the backside helium cooling of the wafer, because there is generally a small leakage rate given. Since its concentration is so low, it can be neglected for the deposition process. However, when the peak is significantly higher than normal, there is likely a problem with the connection between the wafer and the chuck. An observable line is located at $1083,5\,\mathrm{nm}$ (after NIST database [38]).

5.4 Optical spectrum of the cleaning procedure

Regular chamber cleaning is necessary every few wafers to remove $\mathrm{SiO_2}$ from the chamber walls. These cleans are performed automatically by introducing $\mathrm{NF_3}$ into the chamber that is decomposed there into fluorine (F) and nitrogen fluoride (NF). Especially fluorine reacts with the solid silicon dioxide ($\mathrm{SiO_2}$) and forms finally gaseous silicon tetra fluoride ($\mathrm{SiF_4}$) that is removed by the pumping system. In order to accelerate the cleaning the pressure is increased (about a few Torr) in comparison to normal operation where deposition takes place. However, in this pressure regime a plasma generation is not possible, inductively or capacitively, at least with the available equipment that is used for deposition. The heavily reduced mean free path of the free electrons would require much more power to provide decomposition. The gas temperature would be very high, because the plasma would be thermalized so that the gas temperature would be close the electron temperature then. Thus, the excitation is done by microwave radiation by a separate microwave generation unit that directly decomposes the precursors. In that case the gas is not in a plasma state and does virtually not emit light in the optical range. Therefore, a direct monitoring by the OES sensor does not provide useful information. However, the conditioning step at the end of the cleaning procedure, where $\mathrm{SiO_2}$ deposition takes place to condition the chamber wall, can be used for the detection of machine faults before even processing a wafer.

5.5 Time course of the line-intensities

The time course of a few selected line-intensities is depicted in figures 27 and 28. The sampling rate is fixed to one second due the limited capacities of the central data collection system for automatic monitoring of all machine and sensor data for every wafer. As integration time $8\,\mathrm{ms}$ are used, so the spectrum is recorded 125 times per second. The output at every second is then the spectrum averaged over the corresponding 125 recorded spectra.

The first plateau marks the first monitored wafer that is processed by HDP CVD. The long zone (3), where all lines are at zero, indicates the cleaning procedure during which $\mathrm{NF_3}$ is decomposed by microwave radiation at a pressure of a few Torr. The following isolated peak of a few seconds is the result of the conditioning step, that is done to put the chamber in a state that corresponds more or less to the normal state. This minimizes the 'first wafer effect', that is definitely measurable by various process parameters of a wafer that is processed directly after the clean. The subsequent three intensity plateaus that are equal to the first one indicate the next three processed wafers. After them again a cleaning procedure would be recognizable.

Note that the time course of one wafer consists of a main plateau that lasts about 90 seconds (zone 2) and a pre-plateau (zone 1) with a duration of about 10 seconds. While

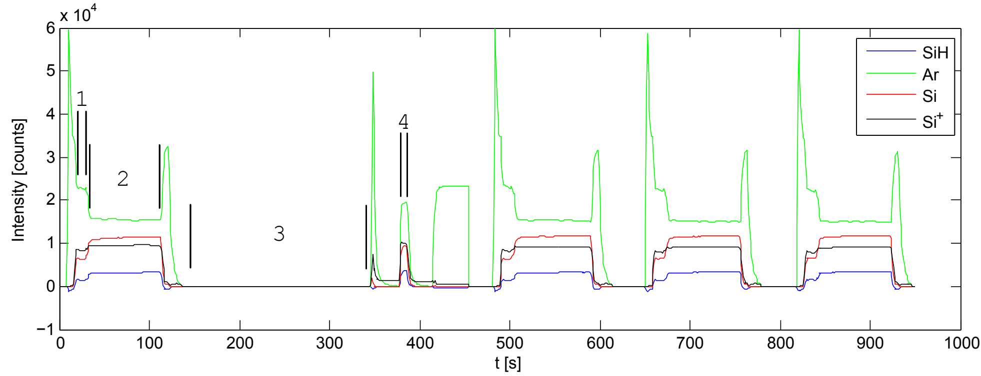


Figure 27: Time-course of SiH, Ar, Si and Si^+ line intensities over 4 wafer and one cleaning cycles. Considering the first wafer (from 0 to about 130 seconds) zone (1) indicates the deposition step without sputtering, while (2) indicates the main deposition step where sputtering takes place. In the following cleaning step (from 140 to about 470 seconds), zone (3) marks the step where NF_3 is injected and zone (4) indicates the conditioning step where deposition takes place to condition the chamber walls.

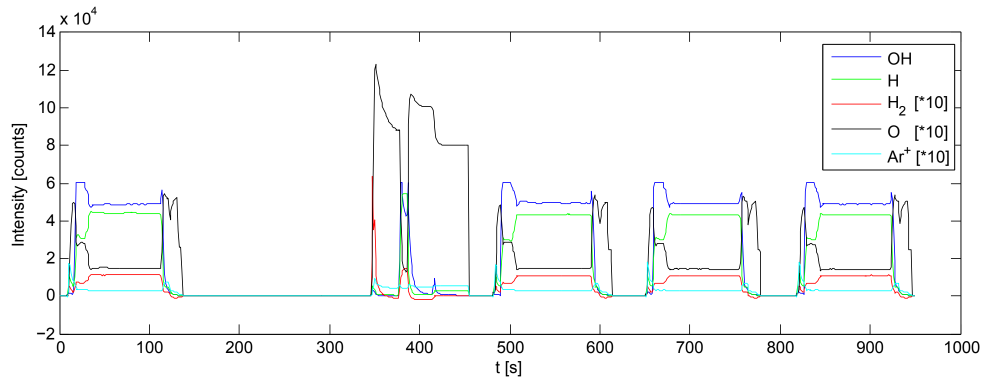


Figure 28: Time-course of the OH, H, H_2, O and Ar^+ line intensities over the same time period as that presented in figure 27.

the main plateau indicates the main deposition step where also sputtering takes place, the shorter plateau is the result of a short pure-deposition step where no sputtering is done. The thereby deposited oxide layer should prevent sensible parts at the wafer surface to the exposure of sputtering. Especially the thin gates would not withstand the impact of high energetic ions.

Since the comparing and evaluation of the process by time courses is rather complicated, indicators are formed by simply taking the average of the line-intensities separately over the pre-plateau and the main plateau. Of course, the same can be done for the conditioning step of the cleaning procedure. As explained in the following chapter also line-intensity-ratios are of interest. In this case the averages are taken of the time course of the line-intensity ratios.

5.6 Determination of some relevant plasma parameters

The main idea of monitoring the HDP CVD process by an optical emission spectrometer is to gain more information about the actual plasma state. Thus, a main question is which information about important plasma parameters can be extracted from the

spectrum. Since for process monitoring absolute values are not required in principle, information about the relative change of certain plasma parameters is actually sufficient. However, the corresponding methods have to be sensitive to small changes of the plasma state, because they are used to be very small in normal operation. Otherwise a prediction of i.e. the deposition rate will hardly be possible.

In fact these requirements are limited not only by the used spectrometer itself, but also because of the nature of the HDP CVD spectrum, since the plasma is a typical example for a technical one. This is not only because of the bunch of plasma species, but also because of the low value of pressure and electron density (in a plasma physical respect) that lead to a non-thermal equilibrium (even local) where electrons have completely different temperatures than ions and neutrals. Moreover, the plasma can not be considered to be optically thin, so self-absorption plays no negligible role, making the determination of absolute plasma parameters much more difficult and inaccurate. Nevertheless, the used spectrometer should provide a high signal-to-noise ratio, a good sensitivity over a broad spectral range as well as a low dark noise (which was actually the reason to prefer the Maya2000 Pro instead of the USB 4000 that has an inadequate signal-noise ratio for the purpose reported here). Additionally, second- and third-order effects should be eliminated by a clever optical filter arrangement. On the other hand only a moderate spectral solution of approx. $1\,\mathrm{nm}$ (or below) is required, because of the broadening of the spectral lines due to the given plasma state.

The most important plasma parameters that can be extracted out of the OES data in principle, are the electron and ion density, the electron temperature, the gas temperature (temperature of neutrals and ions) and the concentration of certain gas phase species. The chamber pressure is provided by the machine itself. As already mentioned, the main objective is an appropriate sensitivity for changes of these parameters, since in the end a statistical model is used for the prediction of interesting process parameters. A purely physical model would be way too complicated and rather inaccurate, because of the bunch of influence parameters, from the plasma as well as the chamber. However, one must bear in mind which informations are necessary for an application of a statistical model and how they have to be prepared for optimization.

5.6.1 Gas temperature

The only possibility to determine the gas temperature with the OES sensor is to investigate the Doppler-broadening of suitable lines. The Doppler-broadening results from the Maxwell-Boltzmann distribution of the particle velocities. However, a better spectrometer resolution ($< 0,1\,\mathrm{nm}$) would be necessary to measure the line profiles accurately enough. But even in this case, the uncertainty (50%-100%) and the resulting poor sensitivity of this procedure would probably not lead to a satisfactory monitoring of the small gas temperature changes at normal operation of the HDP CVD machine during deposition.

5.6.2 Ion density

For this type of plasma, the ion density is assumed to be close to the electron density (compare chapter (3.2.1)). A rough estimation for the ion density gives the line-intensity ratio of two lines of an element like argon and its corresponding ion (Ar^{+}). Changes in the electron temperature can affect this ratio according to section 5.6.4.

5.6.3 Electron density

For its determination there exist a few methods based on OES. The electron density can again be determined by certain line-profiles based on the Stark-effect [44], where the H_α line would be the best choice because of its large broadening. However, the same problems as for the gas temperature would arise by using this method. Additionally the integration time must be shortened extremely to ensure that the intensity of the H_α line (as well as the H_β) is not in the sensor's saturation. But this would cost information about certain species, the line intensities of which are low.

5.6.4 Electron temperature

On the other hand the electron temperature can be determined by the intensity ratio of two lines of the same element, because the excitation energy is different from line to line. According to the Boltzmann factor,

$$e^{-\frac{E}{\mathrm{k_B} T_e}}, \tag{109}$$

the number densities of distinct exited atoms with their respective energy levels depend on the electron temperature T_e. By measuring the line-intensity ratio of two different lines with the excitation energies E_1 and E_2 the electron temperature can then be calculated [44] following,

$$\frac{I_2}{I_1} = \frac{g_2 A_2\, \nu_2}{g_1 A_1\, \nu_1}\, e^{-(E_2 - E_1)/(\mathrm{k_B} T_e)}, \tag{110}$$

where g_1 and g_2 are the statistical weights of the respective levels. ν is the corresponding transition frequency and A the transition probability.

Naturally this method is more sensitive, if the energy difference $(E_2 - E_1)$ is as large as possible. For the two argon lines at 675,3 nm and 703,0 nm mentioned in tabular 2 the corresponding excitation energies are $E_{\mathrm{ArI}\,675} = 14,743\,\mathrm{eV}$ and $E_{\mathrm{ArI}\,703} = 14,839\,\mathrm{eV}$ according to [45]. Since the resulting difference of only about $0,1\,\mathrm{eV}$ is very small (it does not change much for other line pairs), there is also the alternative to use the line of an argon ion as the second line concerning the line intensity ratio. The excitation energy difference is then approx. $5 - 6\,\mathrm{eV}$ [45], mainly because of the different energy gaps between ground state and first excited state. However, in order to obtain the electron temperature the Saha-equation has to be considered, too. Assuming local thermal equilibrium one finds

$$\frac{I^{z+1}}{I^z} = \frac{2(2\pi m_e k_B T_e)^{3/2}}{n_e h^3} \frac{g^{z+1} A^{z+1} \nu^{z+1}}{g^z A^z \nu^z} \exp \frac{-(E^{z+1} - E^z) - E^{z,z+1}}{k_B T_e} \tag{111}$$

where $E^{z,z+1}$ is the ionization energy between the ionization state z and $z+1$. However, one must bear in mind that this equation (111) is only correctly applicable, if the ion density (and with that the electron density) is assumed to be constant (compare chapter 5.6.2). This can not generally be ensured. Moreover, the application of the validity of the Saha-equation presupposes at least a local thermal equilibrium which is definitely not given in this non-equilibrium plasma state. Thus, for monitoring the electron temperature the first method may be the more accurate choice despite its smaller sensitivity. Nevertheless, also this method depends on the electron density, even though not so much.

5.6.5 Simultaneous determination of electron density and electron temperature by line-intensity ratios

Both electron density (n_e) and electron temperature (T_e) influence the line-intensity ratio between two lines of the same element (no matter what charge number) as discussed in chapter 5.6.4 and A. However, the influence varies for n_e and T_e, so that n_e sometimes has a larger influence in comparison to T_e and sometimes a lower influence. This behavior can be exploited by using two line intensity ratios that differ from each other with respect to n_e and T_e. An approach based on this principle - the Collisional Radiative Model (CRM-model) - is discussed in chapter A in detail.

5.6.6 Plasma species concentrations

The probably most important plasma parameters are the concentrations of the gas phase species, since the reactor is likely to be mass transport limited (compare 4.3.6) so that their total amount as well as the proportion to each other is influencing the deposition rate. In order to determine a certain species concentration it is obvious to use a corresponding spectral line for it. However, it could be found out that they are influenced more by other parameters than the actual concentration during normal operation. For example the chamber pressure (compare figure 29) has a considerable influence. Even though, not in the typical way. When the chamber pressure decreases, the line-intensities increase. Usually the line intensities would decrease too. However at this certain non-equilibrium plasma state the line intensities benefit more of the higher energetic electrons accompanied by the lower pressure than of the higher particle density.

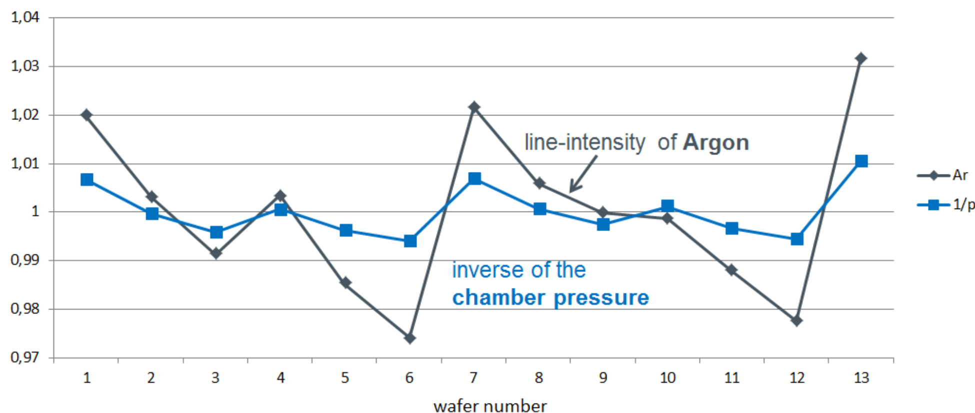


Figure 29: Comparison between the line-intensity of argon and the inverse of the chamber pressure after [46]. Note that both variables were scaled to one.

Additionally, other disruptive effects like a changing electron density or electron temperature, etc. influence the spectral line intensities in a way that they become not representative for the actual relative concentration. Even a simple reason like a slightly replaced optical fiber of the spectrometer can change the total intensity and falsify a model that is based on the intensity of single lines. Thus, it is obvious to use line-intensity ratios rather than single line intensities, especially for determining the proportion of certain gas phase species. As second line ratio a suitable argon line is recommended, since argon is as inert gas hardly affected by any chemical reaction. However, it underlies all other effects that affect the spectral lines of reactants.

5.7 Statistical prediction of the deposition rate

A statistical prediction of the deposition rate based on the OES-sensor makes sense, since it could be shown that the deposition process is mass transport limited (compare 4.3.6). Thus, the concentration of reactants and by-products, represented by line-intensity ratios according to 5.6.6, has a direct influence on the deposition rate. However, there is another important parameter that can not be measured with the OES-sensor, but significantly influences the net deposition rate of the wafer.

5.7.1 Influences of the cleaning cycles

Following a simple thermodynamic classification of the HDP CVD process (compare 2.6.1), the formation of $\mathrm{SiO_2}$ is relieved at lower temperatures mainly because of the accelerated surface reactions. This is confirmed by the work of Meeks et. al [36] where it is shown that the deposition rate tends to be inversely proportional to the substrate temperature. Actually, this is also valid for the chamber wall, where deposition takes place, too (compare 4.3.6). Although the system within the framework of this work is assumed to be mass transport limited, the pure deposition rate (without sputtering) of the wafer and of the chamber wall depends on the ratio of wafer temperature and wall temperature. If the wall temperature increases i.e., the probability that the reactants are consumed decreases and consequently more reactants are available for the consumption at the wafer and vice versa.

Overall, the deposition rate (without sputtering) at the wafer can be influenced by both the surface temperatures of the wafer and of the chamber wall. While the surface temperature of the wafers is assumed to be the same from wafer to wafer at normal operation, this is not mandatory valid for the wall temperature, despite it is tried to keep it constant. Depending on a certain HDP CVD machine type, the chamber has to be cleaned periodically after a fixed amount of wafers (usually after every 2-5th wafer), because on the chamber wall $\mathrm{SiO_2}$ is deposited, too (compare 4.2.5). So the thickness of the $\mathrm{SiO_2}$ layer at the chamber wall depends on the moment when the last cleaning procedure was done. Accordingly, the inner wall temperature increases with an increasing oxide layer, since the $\mathrm{SiO_2}$ layer has a much lower thermal conductivity, which is amplified by the lower density and quality of the $\mathrm{SiO_2}$ layer at the wall in comparison to that on the wafer (because of the lower temperature and the missing sputtering component). Consequently, the deposition rate at the walls decreases until the next cleaning step. On the other hand the deposition rate at the wafer increases, because statistically more reactants are then consumed at the wafer surface (compare figure 33).

Another possible reason for the influence of the last cleaning cycle to the deposition rate is the changing wall surface roughness and with that a change in the relative size of the surface. Since aluminum oxide of which the wall is constructed usually has a higher surface roughness than silicon dioxide, the wall surface roughness decreases the more SiO_2 is deposited on it, leading to a decreased deposition rate at the walls. This would lead again to an increasing deposition rate at the wafer the farther back the last cleaning procedure was done, due to the mass transport limited reactor system.

Anyhow, both possible influence factors - the surface temperatures and the wall surface roughness - may not necessarily be recognized in the optical spectrum, because the amount of reactants within the plasma does not necessarily change. Thus, the

influence of the last cleaning cycle will be taken into account by a simple indicator that is coupled to the cleaning cycles.

5.7.2 Non-linearity of the plasma parameters

Actually, the whole plasma process behaves in a non-linear way, i.a. because of the various electron cross sections of the species. Consequently, the plasma composition is non-linear with respect to the determining plasma parameters (chamber pressure, electron density, electron temperature, ionization degree). Nevertheless, the process is approximately linear within one process step that operates within small boundaries, like every non-linear function that is approximately linear within a range that is small enough, for example the main deposition step (with sputtering) that is used for the prediction of the deposition rate. In order to examine the real behavior of the system concerning non-linearity, both a linear and a non-linear model are used to predict the deposition rate. The non-linear model is based on Gaussian Process Regression (GPR), that is actually a non-parametric model and belongs to the kernel-methods. It does not need any prior information except for a sufficient number of training data. A small introduction to GPR is provided in the appendix B.

5.7.3 Data for modeling

The data available for modeling comprise 53 randomly picked wafers that were processed within 6 weeks. The layer's thicknesses were measured by an ellipsometer and divided by the time interval of the main deposition step to obtain the averaged deposition rate required for modeling. All data were scaled to one. Firstly, because this simplifies the optimization of the hyperparamters of the GPR model a lot. Secondly, the deposition rate is scaled to one because of confidentiality reasons.

It turned out that the combination of the line-intensity ratios $\mathrm{SiH/Ar}$, $\mathrm{OH/Ar}$, $\mathrm{H/Ar}$ and $\mathrm{Si/Ar}$ provide the best statistical prediction of the layer thickness, at least for the sample of wafers. Thus, only these four line-intensity-ratios are used for linear and non-linear modeling. Moreover, also the last cleaning step is considered through an additional variable according to chapter 5.7.1. Since always three wafers are processed until the next clean, this variable can have just three values.

5.7.4 Linear modeling

The result of the linear model applied to the 53 wafers mentioned above is depicted in figure 30. The resulting standard deviation of the predicted deposition rate is $\sigma_{lin} = 0,0086$. From figure 31 it can be seen that there exists approximately a linear relationship.

The R^2 of the linear model is about $78,2\%$, which means in statistic that $78,2\%$ of the variation of the deposition rate is explained by this model. As suspected (compare section 4.4.2), especially the line-intensity ratio of $\mathrm{SiH/Ar}$ tends to be proportional to the deposition rate (compare figure 32, which strongly indicates the predicted mass transport limited deposition process (compare section 4.3.6).

Furthermore, also the suspected influence of the last cleaning cycle could be shown (compare figure 33).

As supposed (compare section 5.7.1), the deposition rate (on the wafer) is significantly smaller directly after the cleaning cycle, since more precursors are consumed

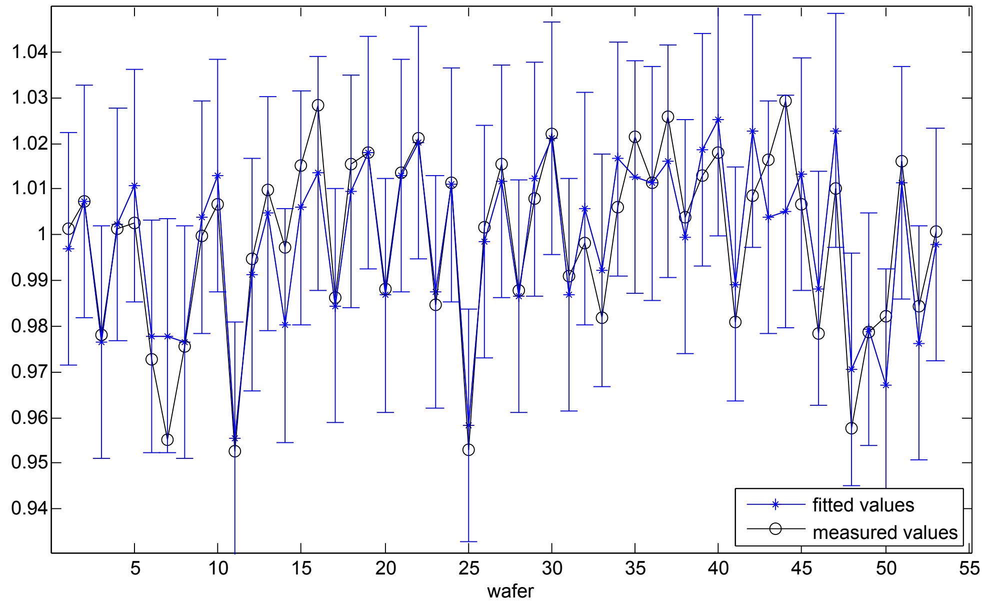


Figure 30: Linear model for prediction of the deposition rate trained on a test sample of 53 wafers. The blue values indicate the fitted deposition rate in comparison to the real measured values in black. Note that the values are scaled to one.

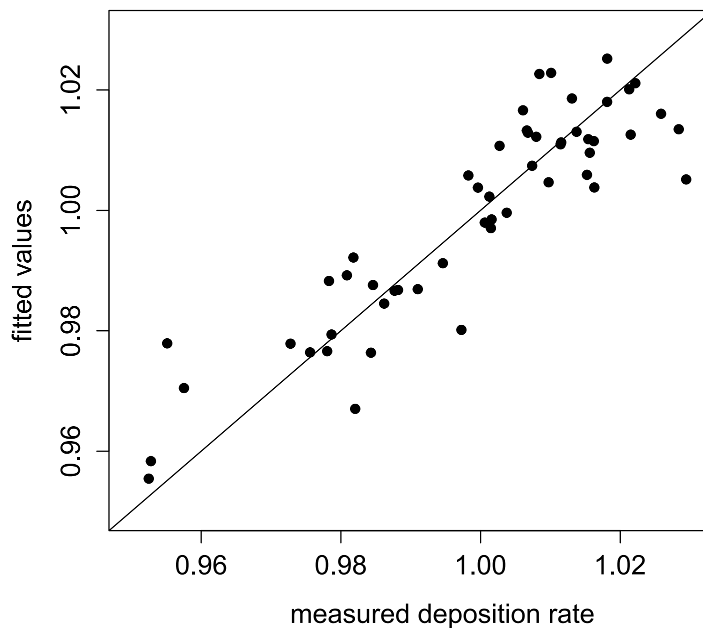


Figure 31: Fitted deposition rates of the linear model versus the real measured deposition rate. If the model would describe the deposition rate without any error, all data points would be on the diagonal line, which would result in an R^2 of 100%. Anyhow, in this certain case the R^2 is at an acceptable value of 78%. The distribution of the fitted values around the diagonal is roughly gaussian. Consequently, the relationship between the input parameters and the deposition rate is approximately linear. However, it can not be excluded that there may exist a non-linear contribution.

at the walls in comparison to wafers that are not processed directly after the cleaning

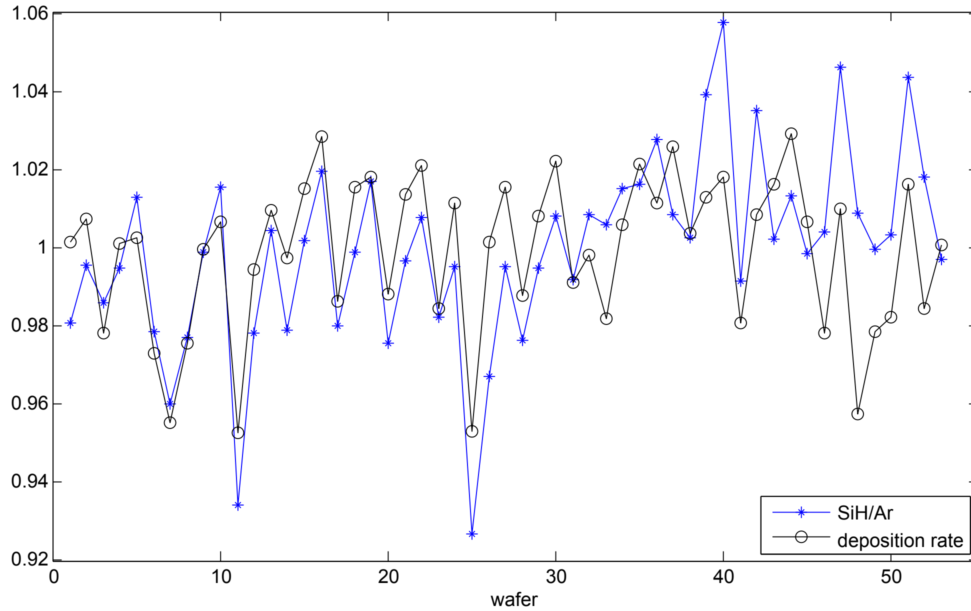


Figure 32: Line-intensity ratio of SiH/Ar in comparison to the deposition rate. For a certain section a mutual proportionality is obvious. In fact, the SiH/Ar line-intensity ratio provides the most information about the actual deposition rate. This is not surprising in view of the surface reactions where SiH_x or SiO is required for the last one that finally leads to surface bounded SiO_2.

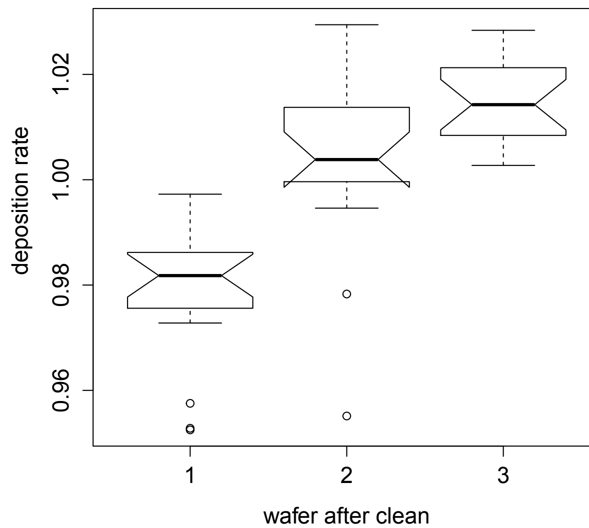


Figure 33: Boxplot of the deposition rate split up according to the last cleaning procedure. 'wafer after clean = 1' means that these wafers are processed directly after the clean. 'wafer after clean = 2' means that these wafers are the second one after the last clean, etc.

cycle.

5.7.5 Non-linear modeling by GPR

The non-linear GPR-model shows quite similar predictions for the deposition rate as the linear model, however, with some improvements. The results of modeling is depicted in figure 34.

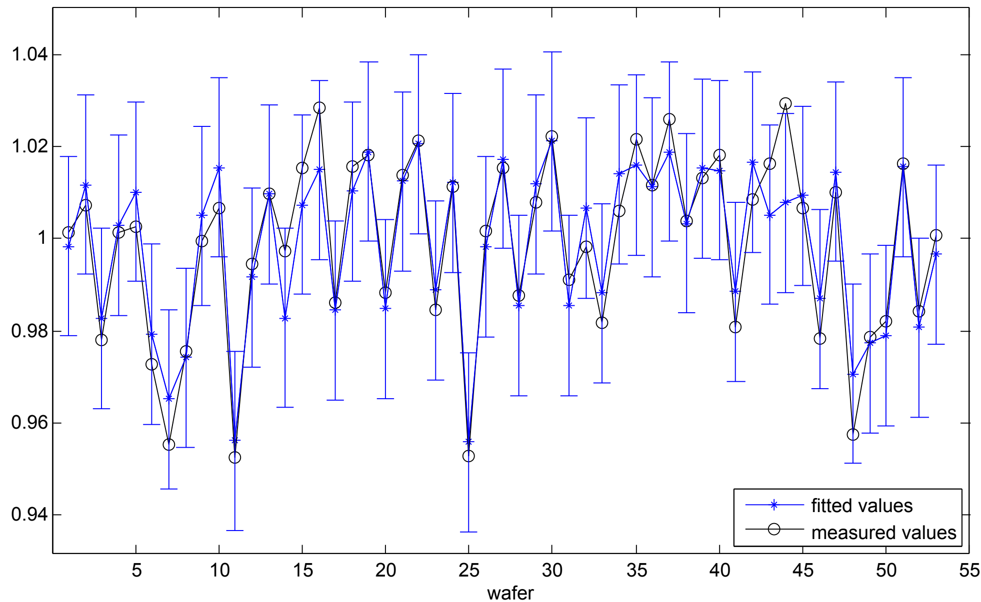


Figure 34: GPR-model for predicting the deposition rate trained on a test sample of 53 wafers. The blue values indicate the therewith fitted deposition rate in comparison to the real measured values in black. Note that the values are scaled to one.

The standard deviation of the predicted deposition rate is now at $\sigma_{\text{GPR}} = 0,0065$, which is about 25% lower than that of the linear model.

5.7.6 Summary

Summarizing it can be stated that the HPD CVD process is very likely to be mass transport limited, because of the direct proportionality of the SiH/Ar line-intensity ratio to the deposition rate. Further, the influence of the cleaning steps could be clearly demonstrated. However, this also states that the available data for modeling may not be complete, since the influence of the last cleaning step actually manifests in a different temperature of the chamber walls. Therefore the surface temperature of the inner chamber wall as well as the surface temperature of the wafer should be also monitored in order to enhance the prediction of the deposition rate. The consideration of the last cleaning step is only a weak substitute in this case. A variation of the temperatures due to other reasons could not be monitored with the available sensors and so its impact to the deposition rate.

The enhancement of the GPR model in comparison to the linear model should not be over-interpreted, since the sample of 53 wafers is not large enough. Long-term experience is not yet available, so that there can be an significant influence of larger maintenance operations. Nevertheless, an R^2 of approx. 80% can be considered as a solid estimation of a well implemented model based on OES-data in combination with the information of the last cleaning procedure. A significant enhancement of the

model's quality should be reached by monitoring the real surface temperatures within the chamber, which is unfortunately not possible with the optical spectrometer.

6 Conclusion

In this master thesis the main task was to explore the benefit of an Optical Emission Spectrometer (OES) for process monitoring of an HDP CVD machine that is used in semiconductor manufacturing for conform deposition of very dense SiO_2 layers on 3d-wafer structures. The first part of the work represents mainly the introduction to this special manufacturing process by starting at the Chemical Vapor Deposition (CVD) processes in general and continuing with a detailed view of the Plasma enhanced chemical vapor deposition (PECVD) processes. The second part of the work analyzes the High Density Plasma CVD (HDP CVD) process and the optical spectrum of it. The focus is put on predicting the net deposition rate on the certain wafers.

A simple transport analysis shows that deposition is either in the starved regime (mass transport limited) or in the differential regime (surface limited) depending on whether deposition caused by surface reactions also takes place on the chamber wall or not. The fact is, that deposition takes definitely place on the chamber wall, which makes regular chamber cleaning with NF_3 necessary. However, it is not clear whether it stems from surface reaction, too. There is no doubt that at least one part comes from the sputtered SiO_2 of the wafer. However, it can be assumed that deposition on the chamber wall based on surface reactions also takes place, since it is accelerated generally at lower temperatures [36]. Unfortunately, about the deposition behavior at that specific wall temperature no precise information are available. Therefore, it can be only assumed with a plausible reason that precursors are also directly consumed at the chamber walls and not only on the wafer, which consequently strongly suggests a mass transport limited deposition process.

In that case, the deposition is limited by the available precursors, so that their concentration is determining the deposition rate. At this point OES can be used, since the available line-intensities through OES provide a very good estimation of the actual concentration of the corresponding gas phase species. Since only its variation is important for a statistical modeling, the line-intensity ratio of a line of the asked species and an argon line is much more suitable than a single line-intensity. This has essentially two reasons. Firstly, it makes the OES-monitoring more immune to any changes in the sensor set-up, because they mostly affect the intensity of the whole spectrum to a similar extent. Secondly, there are physical reasons within the plasma, like a changing chamber pressure, that also mainly affects the total intensity but not necessarily the certain species concentrations. This connection could be proved in section 5.6.6.

A further important factor that influences the deposition rate significantly is the last cleaning cycle (compare figure 5.7.1). Accordingly, the deposition rate increases the farther back the last cleaning procedure was done. It is assumed that this comes from the changing surface temperature of the inner chamber walls. Since the SiO_2 layer has a much lower thermal conductivity than the chamber wall consisting of aluminum oxide, the temperature of the inner surface increases the thicker this SiO_2 layer gets. Consequently, the deposition rate on the chamber wall decreases, so that less precursors are consumed there and more are available for the wafer surface so that the deposition rate increases there, since it is assumed that the process is mass transport limited. Another possible reason for the influence of the cleaning cycle to the deposition rate is the changing wall surface roughness that decreases the more SiO_2 is deposited on it, leading to a decreased deposition rate at the walls and subsequently to a higher one at

the wafer (compare chapter 5.7.1). Since both properties - the surface temperatures and the wall surface roughness - could not be measured within the framework of this work, the consideration of the last cleaning cycle was done in form of an simple indicator.

The monitoring of other important plasma parameters by OES, especially the electron density, the electron temperature and the gas temperature, was not possible in an appropriate way, since the possible procedures are firstly very difficult to perform under these special plasma conditions that are far away from optimum for such analysis. Secondly, they would not provide enough sensitivity for measuring small changes of that parameters because they are mainly based on the corresponding line-profiles so that there is a large uncertainty given. A quite promising procedure based on a Collisional Radiative Model (CRM) that uses line-intensity ratios of different argon lines to determine the electron density and electron temperature simultaneously could also not deliver the desired result. However, because the deposition is mainly affected by the species concentrations that can be measured quite well, possible effects of the electron density, electron temperature and gas temperature, can be indirectly measured to a certain extent.

A direct monitoring of the cleaning procedure is not possible with OES, since the NF_3 gas is decomposed by microwave radiation at a much higher pressure than during deposition. Consequently, there is only extremely low light emission so that the monitoring of the important species is not possible. On the other hand, the potential benefit of OES for Fault Detection and Classification (FDC) can be considered to be very high, since malfunctions of the HPD CVD machine that affect the plasma also affect its spectrum. Unfortunately, no such detectable events occurred within the framework of this work. However, for detecting arcing OES is rather unsuitable, at least in the used form, since such an event takes place usually in the microsecond range. But the minimum integration time of usual OES sensors is in the millisecond range. Here a power sensor that monitors the radio frequency radiation produced by the excitation systems may be the better choice, since a sampling rate in the microsecond is usually no problem in this case. OES is also not so suitable for the detection of particles and impurities, since they simply do not affect the optical spectrum. Thus, a OES can not be used for determining necessary wet cleans and similar maintenance actions that have the task to clean the chamber and to free it from disturbing particles.

It could be shown that with the available sensor data a linear modeling of the prediction rate is possible with an R^2 of about 80% (compare figure 30). An approximately linear dependence could be definitely proved (compare figure 31). Further, a non-parametric GPR-model was applied, that is perfectly suitable for machine learning and the recognition of non-linear relationships. Its result does not differ much of that of the linear model. Nevertheless, it recognized a slightly non-linear behavior, so that its standard deviation is about 25% lower than that of the linear model. However, one must bear in mind, that the test sample of 53 wafers is rather small for a safe statement based on such a powerful model.

However, the results of the linear model can be considered as being representative for linear modeling of the deposition rate of that HDP CVD machine based on the OES-data and the consideration of the last cleaning step. An improvement of the in-situ prediction of the deposition rate can very likely be reached by an in-situ measurement of the surface temperature of both, the just deposited oxide layer of the inner wall and the one of the wafer. Further, also a Fourier Transform Infrared Spectrometer (FTIR)

would very likely enhance the prediction, since with that device the monitoring of a few important gas phase species like SiO, SiH_2, SiH_3, etc. would be possible then.

A Collisional radiative model (CRM)

A.1 Introduction

The Collisional Radiative Model (CR-Model) is a 0-dimensional plasma model, which is normally used to determine the density of the excited states of one or more atomic species as a function of electron temperature and electron density as well as of gas temperature and the densities of the ground state atoms [47]. Since the densities of the excited state atoms correspond to the line intensities, this model can also be used, to determine the electron density and electron temperature with the help of line intensities. Two well-chosen line intensity ratios are sufficient as input parameters. Thus, a critical calibration procedure of the spectrometer for absolute intensity measurements can be circumvented.

The dominating processes, which mainly lead to a change of the electronic population density are electron collisions, radiative processes and de-excitation processes caused by the walls. The assumptions that we make to the model are:

- Molecular species are excluded, so the processes which are responsible for their creation and destruction. A consideration would require serious changes of the model due to vibrational and rotational states [48].
- A Maxwell velocity distribution of the electrons is assumed. This is a questionable assumption discussed in previous papers which propose a high energetic tail in the velocity distribution at the HDP CVD process [49].
- Heavy particle collisions can be neglected in general. So the three-body recombination between two electrons and one ion with the third partner carrying away the resulting momentum. It can be neglected under the assumption that the ionization is up to one percent. Of course the two-body recombination between one electron and one ion can be neglected as well, since it is much less probable in comparison to the three-body recombination because of the missing third partner. Furthermore, the collision of the atoms leading to the ionization of one of them can be neglected in comparison to the dominating processes [50]. However, it is usually considered, because it does not need much effort.

A.2 CRM-model with consideration of the first 14 excited states of argon in an argon plasma

Iordanova and Koleva [50] applied such a model to an argon plasma. It considers the first 14 excited states which are the 1s (Ar^*) and the 2p (Ar^{**}) states , yielding 14 rate equations. The main processes are electron induced transitions from the ground state and the two metastable 1s-states into higher lying states and vice versa. While additionally those kinds of transitions are considered between all four 1s states, they are neglected between the remaining 2p-states. It has also been taken into account that electron impact excitation from the metastable states can lead to ionization. However recombination mechanisms like the three particle recombination ($\mathrm{Ar}^+ + \mathrm{e} + \mathrm{e} \rightarrow \mathrm{Ar}* + \mathrm{e}$) or the less probable (due to the momentum conservation) two particle recombination ($\mathrm{Ar}^+ + \mathrm{e} \rightarrow \mathrm{Ar}* + \mathrm{e}$) are neglected. A further important quantity is the diffusion of the

metastable atoms to the wall, which leads to a de-excitation into the ground state in an inelastic collision. Of course, radiation processes have to be considered for all allowed transitions taking into account reabsorption for some of them as being discussed later. The heavy particle collisions of Ar* atoms leading to ionization have a minor influence on the electronic population, though it is concluded into the CRM-model by Iordanova and Koleva [50]. All considered reaction mechanisms are summarized in the following listings, which also provide the used references for determining the transition coefficients. Ar indicates the ground state, while Ar* and Ar** indicate the four 1s-state and the eight 2p-states.

- Electron impact excitation and de-excitation processes (excitation by inelastic and de-excitation by superelastic electron collision). It corresponds to the terms of the form like **1** and **4** in the subsequent equations (121) to (125)

$$\mathrm{Ar} + \mathrm{e} \longleftrightarrow \mathrm{Ar}^{*} + \mathrm{e} \tag{112}$$

$$\mathrm{Ar} + \mathrm{e} \longleftrightarrow \mathrm{Ar}^{**} + \mathrm{e} \tag{113}$$

$$\mathrm{Ar}_{\mathrm{i}}^{*} + \mathrm{e} \longleftrightarrow \mathrm{Ar}_{\mathrm{j}}^{*} + \mathrm{e} \quad \text{... transitions within the 1s-sates} \tag{114}$$

$$\mathrm{Ar}^{*} + \mathrm{e} \longleftrightarrow \mathrm{Ar}^{**} + \mathrm{e} \tag{115}$$

- Electron induced ionization from the excited 1s and 2p-states. Corresponds to term **3** and its equivalents.

$$\mathrm{Ar}^{*} + \mathrm{e} \longrightarrow \mathrm{Ar}^{+} + \mathrm{e} \tag{116}$$

$$\mathrm{Ar}^{**} + \mathrm{e} \longrightarrow \mathrm{Ar}^{+} + \mathrm{e} \tag{117}$$

- Radiative decay of the Ar* and Ar** states. Corresponds to term **2** and its equivalents.

$$\mathrm{Ar}^{*} \longrightarrow \mathrm{Ar} + \gamma \tag{118}$$

$$\mathrm{Ar}^{**} \longrightarrow \mathrm{Ar} + \gamma \tag{119}$$

- Diffusion of the metastables to the walls with characteristic time $\tau_{1,3}$ leading to de-excitation into the ground state. Corresponds to term **2** and its equivalents.
- Heavy particle collisions leading to ionization. Corresponds to term **6** and its equivalents.

$$\mathrm{Ar}_{\mathrm{i}}^{*} + \mathrm{Ar}_{\mathrm{j}}^{*} \longrightarrow \mathrm{Ar} + \mathrm{Ar}^{+} + \mathrm{e} \tag{120}$$

Consequently the 14 rate balance equations are

$$\overbrace{\sum_{\substack{j=0\\j\neq 1}}^{14} \alpha_{j1} N_j n_e}^{1.} + \overbrace{\sum_{\substack{j=5\\j\neq 10}}^{13} A_{j1} N_j}^{2.} = \left[\left(\overbrace{\alpha_{1i}}^{3} + \overbrace{\sum_{\substack{j=0\\j\neq 1}}^{14} \alpha_{1j}}^{4.}\right) n_e + \overbrace{\tau_1^{-1}}^{5.} + \overbrace{\sum_{j=1}^{4} \alpha_{1j} N_j}^{6.}\right] N_1, \tag{121}$$

$$\sum_{\substack{j=0\\j\neq 2}}^{4} \alpha_{j2} N_j n_e + \sum_{\substack{j=5\\j\neq 6}}^{14} A_{j2} N_j = \left[\left(\alpha_{2i} + \sum_{\substack{j=0\\j\neq 2}}^{4} \alpha_{2j}\right) n_e + \overbrace{A_{20}^{eff}}^{7.} + \sum_{j=1}^{4} \alpha_{2j} N_j\right] N_2, \tag{122}$$

$$\sum_{\substack{j=0\\j\neq 3}}^{14}\alpha_{j3}N_j n_e+\sum_{j=5,8,11,13}A_{j3}N_j=\left[\left(\alpha_{3i}+\sum_{\substack{j=0\\j\neq 3}}^{14}\alpha_{3j}\right)n_e+\tau_3^{-1}+\sum_{j=1}^{4}\alpha_{3j}N_j\right]N_3, \quad (123)$$

$$\sum_{\substack{j=0\\j\neq 4}}^{4}\alpha_{j4}N_j n_e+\sum_{\substack{j=5\\j\neq 6,10}}^{14}A_{j4}N_j=\left[\left(\alpha_{4i}+\sum_{j=0}^{3}\alpha_{4j}\right)n_e+A_{40}^{eff}+\sum_{j=1}^{4}\alpha_{4j}N_j\right]N_4, \quad (124)$$

$$\sum_{j=0,1,3}\alpha_{jm}N_j n_e=\left[\left(\alpha_{mi}+\sum_{j=0,1,3}\alpha_{mj}\right)n_e+\sum_{j<m}A_{mj}^{eff}\right]N_m, \quad m=5...14, \quad (125)$$

where $j=0$ corresponds to the ground state, $1\leq j\leq 4$ to four 1s states and $5\leq j\leq 14$ to the eight 2p states. The index i stands for the ionization state. α is the reaction rate coefficient, whereas A is the radiative transition probability (Einstein coefficient). N_j is the population density of argon in the corresponding level.

While the left side indicates the population processes, the right side represents the depopulation processes. Equations (121) and (123) stand for the two metastable 1s states and consider therefore the diffusion of the metastables to the wall and their subsequent de-excitation expressed by the characteristic time $\tau_{1,3}$ (term 4.) which can be calculate using

$$\tau_{1,3}=\frac{\Lambda}{D_{1,3}}, \quad \Lambda=R/2,405, \quad D_{1,3}=D_{1,3}^{SC}\frac{N_{1,3}^{sc}}{N_0}\sqrt{\frac{T_g}{T_g^{sc}}}, \quad (126)$$

where $D_{1,3}$ are the diffusion coefficients of the two metastables expressed as relation to the standard diffusion coefficient $D_{1,3}^{sc}$ at a temperature of $T_g^{sc}=300\,\mathrm{K}$ and a pressure of $p^{sc}=133,3\,\mathrm{Pa}$. For $D_{1,3}^{sc}N_{1,3}^{sc}$ can be used $D_1^{sc}N_1^{sc}=1,8*10^{18}\,\mathrm{cm^{-1}s^{-1}}$ and $D_3^{sc}N_3^{sc}=1,9*10^{18}\,\mathrm{cm^{-1}s^{-1}}$ [50].

The reaction rate coefficient α_{xi} for electron collisions leading to ionization is [51]

$$\alpha_{ji}=\frac{9,56*10^{-12}(\mathrm{k_B}T_e)^{-3/2}e^{-\frac{\Delta E_j}{\mathrm{k_B}T_e}}}{\left(\frac{\Delta E_j}{\mathrm{k_B}T_e}\right)^{2,33}+4,38\left(\frac{\Delta E_j}{\mathrm{k_B}T_e}\right)^{1,72}+1,32\left(\frac{\Delta E_j}{\mathrm{k_B}T_e}\right)}, \quad 1\leq j\leq 14, \quad (127)$$

where ΔE is the energy difference between the excited state j and the ionization state that is at $15,76eV$ for argon.

As already mentioned re-absorption has to be included for certain transitions because of its influence to the line-intensity ratio measurements. In this case this considered in form of an escape factor assuming an infinite long cylindrical plasma rod with the radius R [52] which reduces the Einstein coefficient A_{21} to an effective transition probability with 2 indicating the upper state and 1 the lower state,

$$A^{eff}=A_{21}\,\eta, \quad \eta=\frac{1,60}{k_0R(\pi\log(k_0R))^{1/2}}, \quad k_0R\gg 1. \quad (128)$$

The absorption coefficient k_0 is

$$k_0 = \frac{A_{21}\lambda^3}{8\pi}\frac{g_2}{g_1}\sqrt{\frac{M^{1/2}}{(2\pi k_B T_g)}}N_1, \tag{129}$$

where N_1 is the population density of the lower state. g_1 and g_2 are the statistical weights of the lower and upper state, while A_{21} is the transition probability (Einstein coefficient) from the lower to the upper state. M and T_g represent the mass of the argon atom and the gas temperature or more precisely the temperature of the non-ionized argon atoms, respectively.

Since η is a strictly monotonically falling function by k_0R (compare figure 35) and gets therefore greater than one for values of k_0R smaller than $1,6$, Iordanova and Koleva [50] assumed that for those transitions η is one,

$$A^{eff} = A_{21}\eta, \quad \eta = \max\left\{1, \frac{1,60}{k_0R(\pi\log(k_0R))^{1/2}}\right\}. \tag{130}$$

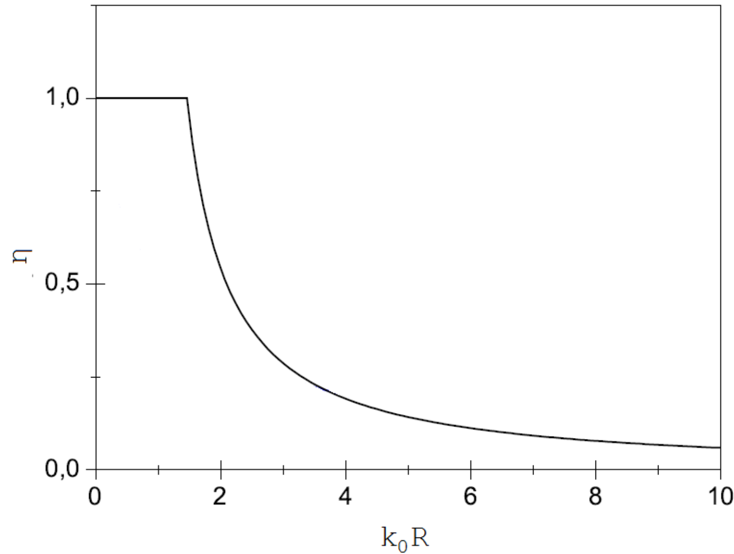


Figure 35: Escape factor η after Iordanova [50]

This raises some questions of its applicability to the CRM-model of Iordanova and Koleva [50] under the given conditions, because the plasma is in a region where k_0R is not necessarily very small or much greater than one. So the transitions in the border region of equation (130) are making serious problems and it is somehow questionable if this is an appropriate approach. Especially equation (130) seems physically pretty implausible, since η drops suddenly very strong when reaching a certain population density.

As indicated in the equations (122) and (124) re-absorption is included for the resonance transitions of the ground state to the two metastable s-states due to the high population density of the ground state [36]. Additionally re-absorption is included for the transition of the metastable 1s-states to the 2p-states, also because of the relatively high population density of the lower 1s-states. But it is restricted [50] to transitions with a corresponding oscillator strength of $f_{12} > 0,119$ which leads us back to the problem with the physically implausible border region behaviour of the η function.

The radiative transition probabilities as well as the statistical weights, excitation energies and wavelengths can be otained from the NIST database [6]. The energy dependent electron impact cross sections are treated theoretically from Bartschat and Zeman [53]. However, the one for excitation from the ground state are likely too large for intermediate energies [54, 55]. Anyway, for the calculation of the corresponding reaction rate coefficients it is necessary to make an assumption to the electron energy distribution. Usually the Maxwellian distribution function is applied. However, this choice is questionable for the HDP CVD plasma, since there is presumably a high energetic tail in the electron energy distribution [49].

The determination of the population density of the ground state N_0 was performed only over the ideal gas law in the original work [50], which was actually modified later by Crintea et al. [56]. It was claimed that there is an additional mechanism of neutral depletion called electron pressure, which has already been treated in O'Connell et al. [57] and O'Connell et al. [58]. The electron pressure only becomes important at a low gas pressure in combination with the high electron density which is characteristic for this type of plasma. The population density of the ground state N_0 can then be expressed as

$$N_g = \frac{p}{k_B T_g} - n_e \left(1 + \frac{T_e}{T_g}\right), \qquad (131)$$

assuming that the ion temperature is equal to the neutral's temperature. This is actually completely incorrect for the HDP CVD plasma, since the ion temperature is increased strongly by the capacitively coupled RF-generator operated at $13,56\,\mathrm{MHz}$ to provide sputtering. However, it can be suggested that the resulting error is modest because of the very small ionization (lower than 1 percent [36]) in the HDP CVD plasma.

Actually this specific CRM-model of Iordanova and Koleva [50] comprises 19 variables, since there are 14 population densities of the included excited argon states as well as the population density of the ground state which make 15 variables. Additionally there is the electron density as well as the electron temperature which also occurs beside equation (131) implicitly in the reaction rate coefficients. The last two remaining variables are the chamber pressure p and the gas temperature T_g. So overall there are 15 equations in comparison to 19 variables. In fact the chamber pressure p can be easily measured in-situ by the sensors of the HDP CVD plant itself, while the gas temperature T_g has to be measured in a more complicated way, since there are normally no sensors scheduled for that. Alternatively T_g can be estimated by a previous theoretical work [36], since it mainly depends on the applied power and the chamber pressure [56] which can both be monitored by the sensors of the machine. So if there are no significant changes in these two parameters the idea is to treat T_g as a constant. Nevertheless it should be kept in mind that the gas temperature has a strong impact not only on the population density of the ground state, but also on the escape factor and the diffusion coefficient of the metastables.

The two remaining equations for the determination of the 19 variables are obtained by two line-intensity ratios. In fact the line-intensity ratios can be obtained easily by knowing the excited level population and vice versa. However, it is important to choose the two line-intensity ratios in such a way that their dependence on the electron density and electron temperature is different as much as possible. The optimal

[6] http://physics.nist.gov/PhysRefData/ASD/lines_form.html

situation for the sensitivity of the method would be that one of the line-intensity ratio depends only on the electron density as well as the other line-intensity ratio depends only on the electron temperature. Anyway, to obtain a sufficient dependence of one line-intensity ratio on the electron temperature or the electron density it is usually necessary to choose the spectral lines so that the energy-dependence of the respective intensity ratios is different [50]. Otherwise the corresponding line-intensity ratio would have virtually no sensitivity to any of the two mentioned parameters.

A.3 Application to a pure argon plasma

As two suitable line-intensity ratios (r_1 and r_2) the lines at $696,5\,\mathrm{nm}$ and $750,4\,\mathrm{nm}$ as well as the lines at $801,5\,\mathrm{nm}$ and $794,8\,\mathrm{nm}$ are suggested [50]. If $r_1 = I_{696,5}/I_{750,4}$ and $r_2 = I_{801,5}/I_{794,8}$ are depicted in a (T_e,n_e) grid of isolines based on the CRM-model for various line-intensity ratios, their cross-points ultimately determine the corresponding values of T_e and n_e. The isolines originally calculated by Iordanova and Koleva [50] are shown in figure 36. The two line-intensity ratios show a quite different behavior with respect to changes of T_e und n_e which results in pronounced cross-points. However, re-absorption of the 1s-2p transition is not considered in these calculations as well as the electron pressure (131). By their consideration the isolines would take a shape like presented in figure 37 which looks completely different, in fact because of both mentioned effects [59]. Every electron temperature two electron densities can be assigned, due to the reversal point which is actually caused by the increasing influence of the electron pressure with increasing electron density.

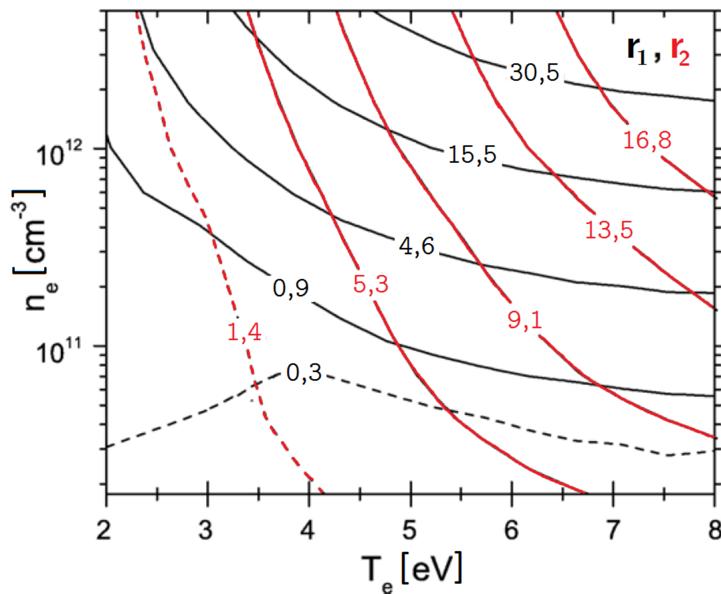


Figure 36: Isolines of the two line-intensity-ratios r_1 and r_2 in the (n_e,T_e) after Iordanova [50] where re-absorption and the effect of electron pressure is supposed [59] to not be considered.

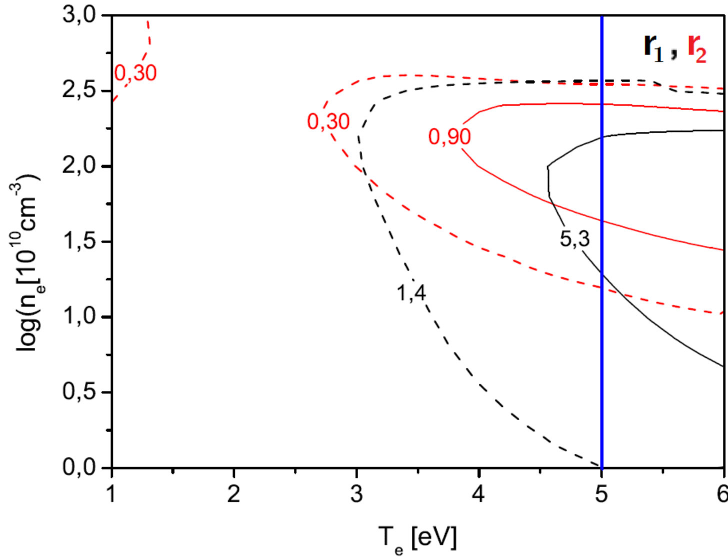


Figure 37: Isolines of the two line-intensity-ratios r_1 and r_2 in the (n_e,T_e)-plane for an ICP argon plasma following the modified CRM-model of Iordanova [50] by Crintea [56]. The figure is generated by Siepa [59]. Here the re-absorption term is included plus the electron pressure that comes from the modification by Crintea [56]. Note that because of the reversal points for one electron temperature two electron densities can be found. Moreover, at the reversal points this method would be very insensitive. A determination of the electron temperature and electron density by the two line-intensity ratios r_1 and r_2 at the HDP CVD would then probably very ineffective even if the model could be applied without further modifications.

A.4 Application to the HDP CVD plasma

Hence in the HDP CVD plasma not only argon is contained the above CRM-model has to be modified. Fortunately, argon does not virtually interact with the other bunch of species, since it is a noble gas. Even under these special plasma conditions there are no significant reaction paths [36]. Nonetheless, there are a few effects that have to be kept in mind. Firstly, the additional heavy particle collisions between argon atoms and other species that can lead to excitation, de-excitation or ionization. However, since the corresponding term in the rate equations for argon-argon collisions already have a minor influence, this can be assumed also for any additional heavy particle collisions. Secondly, there is maybe a significant absorption of photons that are emitted by other species, in certain argon states. This would then lead to a significant distortion of the distribution of the 14 argon states which is not considered in the CRM-model.

Thirdly, the electron density is increased by the additional ions from the other species. However, this does not presumably make difficulties, since it is actually the task of the CRM-model to determine it. Finally, the argon pressure corresponds no longer the chamber pressure because of the other species. In fact, it is necessary to use the partial pressure of argon to calculate the density of argon atoms in the ground state. In order to calculate its proportion in the HDP CVD plasma it has to be determined. One possibility to calculate it, is the Aurora model of E. Meeks et al. [36], which predicts the steady-state properties of a perfectly stirred plasma reactor. It was originally developed for the application to thermal CVD reactor systems [37]. Alterna-

tively, the results of the work of Meeks et al. [36] in which Aurora was applied to the HDP CVD model can be used for an approximation. This case corresponds quite well to the used HDP CVD reactor within the framework of this work and gives a molar proportion of 25,5$\%$ of argon in the plasma, which corresponds to a partial pressure of $p_{\mathrm{Ar}} = 1,72\,\mathrm{mTorr}$ assuming a chamber pressure of $p = 6,75\,\mathrm{mTorr}$ (compare 4.3).

Apart from these problems the reversal points of the isolines, caused by the increasing influence of the electron pressure at low gas pressures and high electron densities, can also prevent or at least hinder a determination of the electron density and temperature with this CRM-model. Thus, for clarity about the applicability of this CRM to the HDP CVD process, experiments where the electron temperature and density are measured should be done. Anyhow, an experiment where the input power of the HDP CVD reactor was varied for about 20$\%$, which should affect the electron density and temperature, didn't show any reproducible changes of the two line-intensity ratios r_1 and r_2. So a suitable application of this model is very unlikely at least with the two line-intensity ratios r_1 and r_2.

B Gaussian process regression (GPR)

The following chapter is only a short introduction into the complex topic of the Gaussian Process Regression. It is mainly oriented on the paper of Rasmussen [60].

A typical prediction problem has the problem that the observations of a dependent variable is noisy. Therefore it is not valid to interconnect the observation points by the smoothest possible curve and consider it as the function between the dependent variable $f(x)$ and the independent variables x. The estimation of the dependent variable at a new value $x*$ would not likely to be accurate. A sound estimation would require the knowledge of the underlying function or at least a good assumption of its properties. But for complex and not completely understood systems this can be very difficult or impossible, respectively. Therefore a non-parametric model like gaussian process regression is an appropriate approach [60]. In comparison to the parametric models they require no knowledge about the underlying function and are actually able to determine them in an oblique way by a simultaneous determination of the overlapping noise. The advantage of gaussian process regression to other typical non-parametric models like artificial neural networks is that it is easier with respect to interpretation and applicability, whereby it is suitable for supervised machine learning (another non-parametric model which has similar properties is the support vector machine).

A gaussian process can be considered as a generalized gaussian distribution extended to infinitely many random variables of which any finite subset of the range follows a multivariate gaussian distribution. Similar to a multivariate gaussian distribution $X \approx N_P(\mu, \Sigma)$ which is specified by its mean vector and covariance matrix the gaussian process $f(x) \approx GP(m, k)$ is fully specified by its mean function $m(x)$ and its covariance function $k(x, x')$ [60]. Whereas the multivariate gaussian distribution is a gaussian distribution of vectors, the gaussian process is a gaussian distribution of functions. In comparison to a gaussian distribution a gaussian process can not be written in an analytical way. However, it is possible for the covariance function $k(x, x')$.

The covariance function $k(x, x')$ is also known as kernel-function, hence the GP is also known as kernel-method. With the help of the kernel-function the kernel-methods enable the computing of inner products between the images of all pairs of data in the so-called feature space. Thus, an explicit transformation of the coordinates of the individual data points in that space is thereby not necessary and saves computational effort. The approach to replace the inner products by the kernel-function is called the kernel-trick. In principle every positive definite function can be used as kernel-function. However, the kernel-function is mostly chosen from a small pool of common functions [61]. In our case we will use the RBF-kernel (Radial Basis Function),

$$k(x, x') = \sigma_y^2 \, e^{-\frac{(x-x')^2}{2l^2}}, \tag{132}$$

with the two hyperparameters σ_y and l. l is called the specific length scale and contributes significantly to the complexity of the model, as it is discussed later. In order to take noise into account, a gaussian distribution term is added for f(x): $y = f(x) + N(0, \sigma_n^2)$. This can be easily implemented by a diagonal matrix with the value σ_n^2 which is added to the covariance matrix,

$$k(x, x') = \sigma_y^2 \, e^{-\frac{(x-x')^2}{2l^2}} + \sigma_n^2 \, \delta_{ii'}. \tag{133}$$

However, if the data are prone to outliers the GP-model can be tuned more robust by using a more heavy-tail noise distribution such as a Laplace or Student-t-distribution [61]. Alternatively there are so-called mixture noise models which provide robustness by separate noise models for regular and extreme observations respectively [62].

In order to train the GP model it is necessary to adjust the mean $m(x)$ and covariance function $k(x, x')$ suitable for the trainings data. For that it is useful to parameterize the functions in terms of hyperparameters as it is already done for the covariance function in form of the RBF-kernel [60],

$$y(x) \approx GP(m, k), \quad m(x) = ax^2 + bx + c, \quad k(x, x') = \sigma_y^2 \, e^{-\frac{(x-x')^2}{2l^2}} + \sigma_n^2 \, \delta_{ii'}, \tag{134}$$

with the hyperparameters $\Theta = a, b, c, \sigma_y, l, \sigma_n$ assuming that the mean function m(x) is not of a higher order than two. In order to choose the best composition of hyperparameters θ for the GP-model, a maximation of the a-posteriori-estimate of θ is necessary, $max\,\{\mathbf{p}(\theta|\mathbf{x}, \mathbf{y})\}$. According to Bayes' theorem, $\mathbf{p}(\theta|\mathbf{x}, \mathbf{y})$ corresponds to $\mathbf{p}(y|\mathbf{x}, \theta)$, which is provided by [63],

$$\log p(y|x, \theta) = -\frac{1}{2}\,\log|K| - \frac{1}{2}\,(\mathbf{y} - \mathbf{m(x)})^T\; K^{-1}\;(\mathbf{y} - \mathbf{m(x)}) - \frac{n}{2}\,\log(2\pi) \tag{135}$$

An optimization of that equation on the basis of θ provides the best set of hyperparameters. However, it must be taken care that for non-trivial gaussian processes the optimization might not be a convex problem and therefore contains local extrema.

The first term $-\frac{1}{2}\log|K|$ measures the complexity of the GP-model and penalizes it accordingly whereas the second term $\frac{1}{2}(\mathbf{y} - \boldsymbol{\mu})^T K^{-1}(\mathbf{y} - \boldsymbol{\mu})$ penalizes the model for its discrepancy between the fitted function and the training data outputs. This provides an automatic trade-off between penalty and data-fit, without a manual setting of the model's complexity. In fact it is easily possible to fit the training data exactly by setting the noise level σ_n to zero. Nevertheless, it would probably not correspond to the real behaviour of the system. The third term $\frac{n}{2}log(2\pi)$ results from the normalization of the marginal likelihood and play therefore no specific role [60].

Once the hyperparameters are optimized based on the trainings data $y(\mathbf{x})$ the GP-model can be used for the estimation of the test data output $y(\boldsymbol{x_*})$. Therefore we look at the joint Gaussian distribution,

$$\begin{bmatrix} y \\ y_* \end{bmatrix} \approx N\left(\begin{bmatrix} m \\ m_* \end{bmatrix}, \begin{bmatrix} K & K_* \\ K_*^T & K_{**} \end{bmatrix}\right), \tag{136}$$

where K is the training data covariance, K_* the covariance between training and test data and K_{**} the test data covariance. The conditional distribution of y_* given y is, based on the formula for conditioning, a joint gaussian distribution,

$$y_*|y \approx N\left(m(x_*) + K_*^T K^{-1}(y - m(x)), K_{**} - K_*^T K^{-1} K_*\right). \tag{137}$$

Consequently, the corresponding conditional distribution of y_*, given y, of a gaussian process is

$$y_* = m(x_*) + K_*^T K^{-1}(y - m(x)) \tag{138}$$

$$var(y_*) = K_{**} - K_*^T K^{-1} K_* + \sigma_n^2 \delta_{ii'}, \tag{139}$$

with [63]

$$K = \begin{bmatrix} k(x_1, x_1) & k(x_1, x_2) & \dots & k(x_1, x_n) \\ k(x_2, x_1) & k(x_2, x_2) & \dots & k(x_2, x_n) \\ \vdots & \vdots & \ddots & \vdots \\ k(x_n, x_1) & k(x_n, x_2) & \dots & k(x_n, x_n) \end{bmatrix} \quad (140)$$

$$K_* = \begin{bmatrix} k(x_*, x_1) & k(x_*, x_2) & \dots & k(x_*, x_n) \end{bmatrix} \quad (141)$$

$$K_{**} = k(x_*, x_*). \quad (142)$$

Note that it is mostly possible to set the mean-function zero, because the covariance-function is actually sufficient to describe the relations between one observation to another [60]. The equations and calculations get therefore easier and moreover the amount of hyperparameters is restricted to three provided that a RBF-kernel is used.

The following figures [7] should demonstrate to influence of the length scale l as hyperparameter to the properties of the gaussian process and what an incorrect optimized set of hyperparameters can cause with respect to the data fit, respectively. Note that the data points in the three figures are always the same.

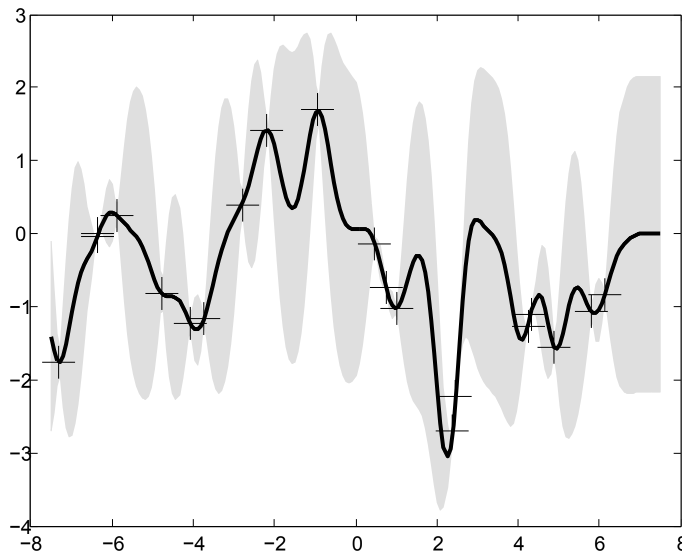


Figure 38: GPR with hyperparameter optimized under the assumption $l = 0.3$. The noise σ_n^2 gets virtually zero and the predicted mean interpolates almost exactly the data points. Note that the error bars grow rapidly if they get further away of the data points, which is indeed due to the small length scale. Consequently the model is very likely to be too complex or at least there are way to few trainings data provided for a sound model building

[7]generated with the gaussian process regression and classification' Matlab scripts of C. E. Rasmussen and C. William (copyright (c) 2005-2007) which is available from http://www.gaussianprocess.org/gpml/code

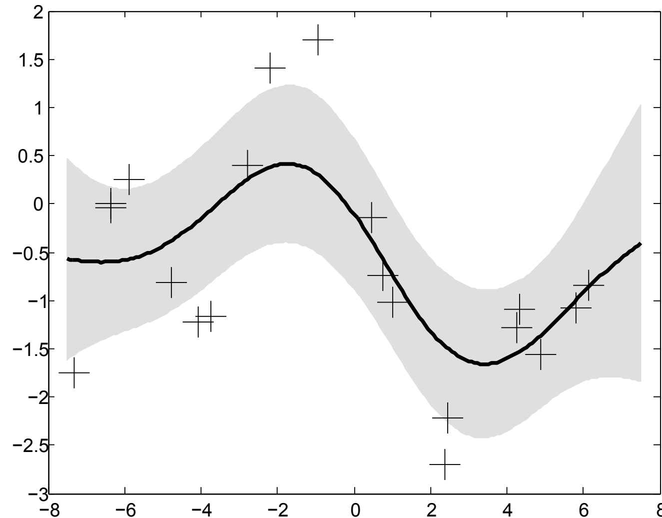


Figure 39: GPR with hyperparameter optimized under the assumption $l = 3$. This is the other extreme case. The fitted function is obviously too simple and leads therefore to a high noise. The error bars do not change much if they get further away from the data points because the length scale is now ten times longer.

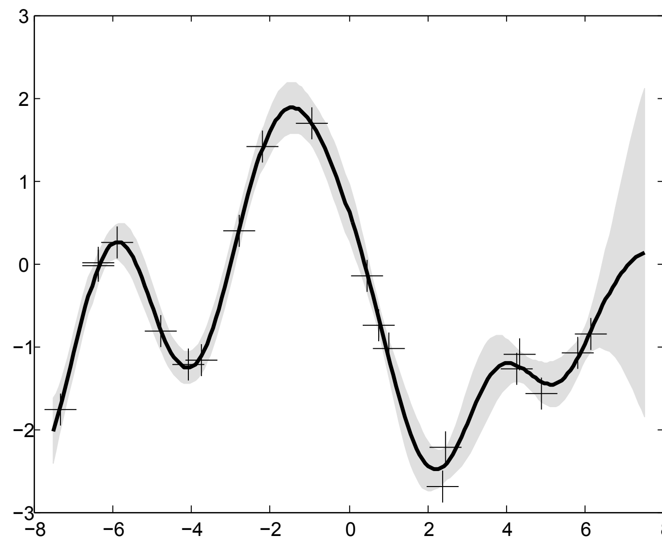


Figure 40: GPR with all hyperparameters optimized with no assumption of l. It can easily be seen that the fitted function by the GP-model is very likely to be the right one. In addition, the confidence interval is very tight throughout the trainings data and therefore the number of trainings data turns out to be sufficient for a sound interpolation. However, a comprising extrapolation makes problems as it can be seen at the edges of the figure. This is especially the case if the underlying function is more complex.

If a GP-model is trained on a multi-dimensional problem with input dimensions that are highly different in mean and standard deviation, it may be difficult to find the global minimum for the hyperparameters. In this case the problem should be bypassed by rescaling the input variables.

Overall the gaussian process regression is a powerful prediction tool which is highly adaptable to also not well understood systems since it doesn't need any prior information except for a sufficient number of trainings data. It is therefore often used for supervised machine learning since the understandability is very high compared to other non-parametric models. A problem may be the computational effort that is of a complexity of $O(n^3)$, since the training of the GP-model requires an inversion of the training covariance matrix K. But there exist approaches for reducing the computational effort by using sparse approximations.

Further information about gaussian process regression can be found under the website
http://www.gaussianprocess.org/ where several links to papers and books are provided. For a good introduction Rasmussen [60] and Ebden [63] are recommended. A more advanced literature is i.e. provided by Rasmussen and Williams [64] and Seeger [61].

References

[1] M. K. Zuraw and D. M. Dobkin. *Principles of Chemical Vapor Deposition*. Kluwer Academic Publishers, 2003.

[2] K. L. Choy. Chemical vapor deposition of coatings. *Progress in Materials Science*, 48:57–170, 2003.

[3] J. R. Creighton and P. Ho. *Chemical Vapor Deposition*, chapter Introduction to Chemical Vapor Deposition (CVD), pages 1–23. ASM International, 2001.

[4] K. Burak-Ucer. Chemical vapor deposition (cont.). `http://users.wfu.edu/ucerkb/Nan242/L10-CVD_b.pdf`, June 2013.

[5] H. O. Pierson. *Handbook of chemical vapor deposition (CVD): Principles, Technology, and Applications*. Noyes Publications, 2 edition, 1999.

[6] Juny 2013. `http://www.timedomaincvd.com/CVD_Fundamentals/Fundamentals_of_CVD.html`.

[7] E. L. Cussler. *Diffusion - Mass Transfer in Fluid Systems*. Cambridge University Press, 1997.

[8] J. Hagen. *Technische Katalyse - Eine Einführung*. VCH, 1996.

[9] Kasper. Heterogene katalyse, vorlesung tc i (sose 2010) teil 2.

[10] Plasma processing, July 2013. `http://www.plasma.inpe.br/LAP_Portal/LAP_Site/Text/Plasma_Processing.htm`.

[11] P. May. Msc physics of advanced semiconductor materials, plasmas and plasma processing, July 2013. `http://www.chm.bris.ac.uk/~paulmay/misc/msc/msc4.htm`.

[12] R. J. M. M. Snijkers. *The sheath of an RF plasma: measurements and simulations of the ion energy distribution*. PhD thesis, Eindhoven University of Technology, 1993.

[13] A. J. van Roosmalen, J. A. G. Baggerman, and S. J. H. Brader. *Dry Etching for VLSI*. Springer, 1991.

[14] B. Chapman. *Glow Discharge Processes - Sputtering and Plasma Etching*. Wiley, 1980.

[15] K. Koehler, J. W. Coburn, D. E. Horne, E. Kay, and J. H. Keller. Plasma potentials of 13.56-mhz rf argon glow discharges in a planar system. *Journal of Applied Physics*, 57(1):59–66, 1985.

[16] R. W. Gould. Radio frequency characteristics of the plasma sheath. *Physics Letters*, 11:236–237, 1964.

[17] J. W. Coburn and E. Kay. Positive-ion bombardment of substrates in rf diode glow discharge sputtering. *Journal of Applied Physics*, 43(12):4965–4971, 1972.

[18] C. B. Zarowin. Plasma etch anisotropy - theory and some verifying experiments relating ion transport, ion energy, and etch profiles. *Journal of the Electrochemical Society*, 130:1144–1152, 1983.

[19] A. Metze, D. W. Ernie, and H. J. Oskam. Application of the physics of plasma ssheath to the modeling of rf plasma reactors. *Journal of Applied Physics*, 60:3081–3087, 1986.

[20] P. M. Vallinga, P. M. Meijer, and F. J. de Hoog. Sheath properties of rf plasmas in a parallel plate etch reactor; the high-frequency regime ($\omega > \omega_i$). *Journal of Physics D: Applied Physics*, 22(11):1650, 1989.

[21] 08 2013. `http://de.wikipedia.org/wiki/Paschen-Gesetz`.

[22] M. A. Lieberman and A. J. Lichtenberg. *Principles of Plasma Discharges and Materials Processing, 2nd Edition*. Wiley, 2005.

[23] M. J. Kushner, W. Z. Collison, M. J. Grapperhaus, J. P. Holland, and M. S. Barnes. A three-dimensional model for inductively coupled plasma etching reactors: Azimuthal symmetry, coil properties, and comparison to experiments. *Journal of Applied Physics*, 80(3):1337–1344, 1996.

[24] V. I. Kolobov and D. J. Economou. The anomalous skin effect in gas discharge plasmas. *Plasma Sources Science and Technology*, 6(2), 1997.

[25] Y. O. Tyshetskiy. *Anomalous and nonlinear eeffect in inductively coupled plasmas*. PhD thesis, Department of Physics & Engineering Physics, University of Saskatchewan, 2003.

[26] V. I. Kolobov, D. P. Lymberopoulos, and D. J. Economou. Electron kinetics and non-joule heating in near-collisionless inductively coupled plasmas. *Physical Review E*, 55(3):3408–3422, 1997.

[27] R. A. Stewart, P. Vitello, D. B. Graves, E. F. Jaeger, and L. A. Berry. Plasma uniformtiy in high-density inductively coupled plasma tools. *Plasma Sources Science and Technology*, 4(1):36–46, 1995.

[28] D. J. Economou, T. J. Bartel, R. S. Wise, and D. P. Lymberopoulos. Two-dimensional direct simulation monte carlo of reactive neutral and ion flow in a high density plasma reactor. *IEEE Transactions on Plasma Science*, 23:581–590, 1995.

[29] G. DiPeso, V. Vahedi, D. W. Hewett, and T. D. Rognlien. Two-dimensional self-consistent fluid simulation of radio frequency inductive sources. *Journal of Vacuum Science & Technology, A: Vacuum, Surfaces, and Films*, 12(4):1387, 1994.

[30] P. L. G. Ventzek, M. Grapperhaus, and M. J. Kushner. Investigation of electron source and ion flux uniformtiy in high plasma density inductively coupled etching tools using two-dimensional modeling. *Journal of Vacuum Science & Technology, B: Microelectronics and Nanometer Structures*, 12(6):3118, 1994.

[31] M. J. Kushner and A. V. Vasenkov. Elsource velocity ddistribution in an inductively coupled plasma. *IEEE Transactions on Plasma Science*, 33(2):388–389, 2005.

[32] K. Iskenderova. *Cleaning Process in High Density Plasma Chemical Vapor Deposition Reactor*. PhD thesis, Drexel University, October 2003.

[33] K. Lai, K. Kumar, A. Chou, and J. C. Lee. Effects of oxide exposure, photoresist and dopant activation on the plasma damage immunity of ultrathin oxides and oxynitrides. In *Electron Devices Meeting, 1995. IEDM '95., International*, pages 319–322, December 1995.

[34] S. Daniels, S. Glynn, F. Soberon, and P. Maguire. Method and apparatus for the detection of arc events during the plasma processing of a wafer, surface of substrate, 02 2011.

[35] E. Yan, J. Xu, X. Li, J. Chen, Y. Chen, D. J. Zhao, L. Chan, L. Ji, D. Padhi, and B. Zhang. Solving arcing issues in cvd processes. *ECS Transactions - The Electrochemical Society*, 18:581–586, 2009.

[36] E. Meeks, R. S. Larson, P. Ho, S. M. Han, E. Edelberg, E. Aydil, and C. Apblett. Modeling of high-density-plasma deposition of SiO_2 in $SiH_4/O_2/Ar$. Technical report, Sandia National Laboratories, 1997.

[37] E. Meeks, H. K. Moffat, J. F. Grcar, and R. J. Kee. Aurora: A fortran program for modeling well stirred plasma and thermal reactors with gas and surface reactions. Technical report, Sandia National Laboratories, 1996.

[38] 2013. `http://www.nist.gov/pml/data/asd.cfm`.

[39] A. Aguilar, J. M. Ajello, R. S. Mangina, and G. K. James. The electron-excited mid-ultraviolet to near-infrared spectrum of h2: Cross sections and transition probabilities. *The Astrophysical Journal Supplement Series*, 177:388–407, 2008.

[40] R. W. B. Pearse and A. G. Gaydon. *The identification of molecular spectra*. Chapman and Hall, 4 edition, 1984.

[41] J. A. Theil and G. Powell. The effects of he plasma interactions with sih4 in remote plasma enhanced chemical vapor deposition. *Journal of Applied Physics*, 75(5):2652–2666, 1994.

[42] K. P. Huber and G. Herzberg. *Molecular Spectra and Molecular Structure, IV. Constants of Diatomic Molecules*. Van Nostrand Reinhold Ltd., 1979.

[43] September 2013. `http://www.astrosurf.com/buil/baader/eval_us.htm`.

[44] R. Rompe and M. Steenbeck. *Ergebnisse der Plasmaphysik und der Gaselektronik*. Akademie-Verlag Berlin, 1967.

[45] M. C. Ötztürk, H. Iwai, D. L. Kwong, P. J. Timans, E. P. Gusev, S. J. Koester, and F. Roozeboom, editors. *Advanced Gate Stack, Source/Drain and Channel Engineering for Si-Based CMOS 3: New Materials, Processes and Equipment*. ESC transactions, 2007.

[46] M. Haselmann. In-situ optical plasma diagnostics in high-density-plasma chemical vapor deposition of sio2 on semiconductor wafers. Dresden, 2013. APCM Conference.

[47] A. Hartgers, J. van Dijk, J. Jonkers, and J. A. M van der Mullen. CRModel: A general collisional radiative modeling code. *Computer Physics Communications*, 135:199–218, 2001.

[48] B. van der Sijde, J. A. M. van der Mullen, and D. C. Schram. Collisional radiative models in plasmas. *Beiträge aus der Plasmaphysik*, 24:447–473, 1984.

[49] H. Li. Measurements of electron energy distribution function and neutral gas temperature in an inductively coupled plasma. Master's thesis, University of Saskatchewan, 2006.

[50] S. Iordanova and I. Koleva. Optical emission spectroscopy diagnostics of inducitvely-driven plasmas in argon gas at low pressures. *Spectrochimica Acta, Part B: Atomic Spectroscopy*, 62:344–356, 2007.

[51] L. Vriens and A. H. M. Smeets. Cross-section and rate formulas for electron-impact ionization, excitation, deexcitation, and total depopulation of excited atoms. *Physical Review A: Atomic, Molecular, and Optical Physics*, 22(3):940–951, 1980.

[52] T. Holstein. Imprisonment of resonance radiation in gases II. *Physical Review*, 83(6):1159–1168, 1951.

[53] K. Bartschat and V. Zeman. Eplasmas-impact excitation from the $(3p^5 4s)$ metastable states of argon. *Physical Review A: Atomic, Molecular, and Optical Physics*, 59(4), 1999.

[54] O. Zatsarinny and K. Bartschat. Large-scale pseudostate calculations for electron scattering from neon atoms. *Physical Review A: Atomic, Molecular, and Optical Physics*, 85(062710), 2012.

[55] O. Zatsarinny and K. Bartschat. Electron-impact excitation of neon at intermediate energies. *Physical Review A: Atomic, Molecular, and Optical Physics*, 86(022717), 2012.

[56] D. L. Crintea, U. Czarnetzki, S. Iordanova, I. Koleva, and D. Luggenhölscher. Plasma diagnostics by optical emission spectroscopy on argon and comparison with thomson scattering. *Journal of Physics D: Applied Physics*, 42(4), 2009.

[57] D. O'Connell, T. Gans, D. L. Crintea, and U. Czarnetzki. Plasma dynamics in an inductively coupled magnetic neutral loop discharge. *Plasma Sources Science and Technology*, 17(2), 2008.

[58] D. O'Connell, T. Gans, D. L. Crintea, U. Czarnetzki, and N. Sadeghi. Neutral gas depletion mechanisms in dense low-temperature argon plasmas. *Journal of Physics D*, 41(3), 2008.

[59] S. L. Siepa. Bestimmung von plasmaparametern über optische spektroskopie an argon-plasmen. Master's thesis, Ruhr-Universität Bochum, 2011.

[60] C. E. Rasmussen. Gaussian processes in machine learning. In J. G. Carbonell and J. Siekmann, editors, *Advanced Lectures on Machine Learning*, Lecture Notes in Computer Science: Lecture Notes in Artificial Intelligence, 3176, pages 63–71. Max Planck Institute for Biological Cybernetics, 72076 Tübingen, Germany, Springer, Germany, 2004.

[61] M. Seeger. Gaussian processes for machine learning. *International Journal of Neural Systems*, 14(2):69–106, 2004.

[62] M. Kuss. *Gaussian Process Models for Robust Regression, Classification, and Reinforcement Learning*. PhD thesis, Technical University of Darmstadt, March 2006.

[63] M. Ebden. Gaussian processes for regression: A quick. http://www.robots.ox.ac.uk/ mebden/reports/GPtutorial.pdf, 2008.

[64] C. E. Rasmussen and C. K. I. Williams. *Gaussian Processes for Machine Learning*. MIT Press, 2006.