universität
wien

Masterarbeit

Titel der Masterarbeit

PBSens:
A mathematical Model for Nanowire Gas-Sensors

Verfasser

Klemens Katterbauer BSc.

angestrebter akademischer Grad

Master of Science (MSc.)

Wien, Dezember 2010

Studienkennzahl lt. Studienblatt:	A 066 821
Studium lt. Studienblatt:	Masterstudium Mathematik
Betreuerin / Betreuer:	PD DI. Dr. Clemens Heitzinger

Acknowledgements

First of all I want to thank all the people that have supported me during my life and studies, in particular:

I want to thank my supervisor Clemens Heitzinger for offering me to work within the WWTF project *Mathematics and Nanosensors* (project No.MA09-028) and for his essential remarks and comments he made during our discussions.

Furthermore, I want to thank Martin Vasicek who helped me with questions about physics at the beginning for his remarks and advice. In addition, I want to thank Stefan Baumgartner, Alena Bulyha, Marina Rehrl and Nathalie Tassotti for the comments and discussions that provided essential feedback. I also want to thank Anton Köck (deputy department head at the Austrian Institute of Technology), Elise Brunet and the other team members for our discussions, their support and in particular their measurement data. Last but not least I want to thank my parents for all their support and understanding during my whole life.

Abstract

Nanowire gas sensors have many applications in health-care, safety, environmental montoring etc., yet the major problem that current research faces is a lack of understanding of the reactions that take place between the sensing film and the gas species and hence a substantial lack of selectivity in SnO_2 nanowire sensing devices.

We will present a mathematical model for SnO_2 nanowire sensors. The surface reactions are described by parameter-dependent ordinary differential equations (ODEs) that give the net exchange of electrons between the gas and the SnO_2 nanowire. The net exchange of electrons is then included in the nonlinear Poisson-Boltzmann equation for the computation of the electric potential. From the electric potential we obtain the concentrations of holes and electrons via Boltzmann distributions and finally the graded channel approximation returns the current.

The parameters of 4 different ODE models for the surface reactions have been determined by comparison of the ODE models with measurement data using inverse-modeling techniques. For each of the models, between 5 and 9 parameters were estimated, while the nonlinear nature of the model complicates inverse modeling. The results from the best parameter sets for each of the 4 models were compared, which also resulted in the affirmation of the hypothesis that chemisorption at SnO_2 nanowires is a slow process compared to the ionization reaction.

The equation corresponding to the affirmed hypothesis performs best compared to other model equations in representing accurately the measurement curve. The components of the parameter set are almost equilibrated and these factors are supported by the results of an F-Test used for the statistical comparison of model equations. The simulation results are always within the range of 5% with respect to the current measurements. Therefore this modeling procedure can yield predictive simulations of field-effect gas sensors. This Master thesis was supported by the WWTF (Viennese Science and Technology Fund) project No.MA09-028.

Zusammenfassung

Nanowire Gassensoren können weitreichend angewandt werden, wie z.B. im Gesundheitsbereich, Sicherheitsbereich, und der Umweltüberwachung. Das größte Problem welches Forscher beschäftigt ist die mangelnde Selektivität und die mangelnde Kenntnis über die Reaktionen welche an der Oberfläche des Gassensors auftreten.

Wir präsentieren in dieser Masterarbeit ein mathematisches Model für einen SnO_2 Nanowire Gassensor. Die Oberflächenreaktionen werden durch eine parameterabhängige gewöhnliche Differentialgleichung (GDG) beschrieben. Diese liefert den Nettoaustausch zwischen dem auftretenden Gas und dem SnO_2 Nanowire. Dieser Nettoaustausch wird dann als Inputgröße im nichtlinearen Poisson-Boltzmann Modell verwendet. Vom berechneten elektrischen Potential erhalten wir über die Boltzmannverteilung die Konzentrationen der Löcher und Elektronen und schlussendlich liefert die Graded Channel Approximation die Stromstärke I.

Die Parameter der 4 verschiedenen GDG Modellen für die Oberflächenreaktionen wurden bestimmt durch einen Vergleich der Modelle mit den Messdaten mithilfe von Inverser Modellierungstechniken. Für jedes Modell wurden zwischen 5 und 9 Parameter geschätzt, wobei die Nichtlinearität der Gleichungen den Schätzprozess verkompliziert. The Resultate von den besten Parametersätzen wurden verglichen, welches zur einer Bestätigung der Hypothese dass die Chemisorption auf SnO_2Nanowires ein langsamer Prozess verglichen mit der Ionisierungsreaktion ist.

Die Gleichung der bestätigten Hypothese beschreibt verglichen mit den anderen Modellen am Besten die Messkurve. Die Komponenten des Parametersatzes sind fast von gleicher Größe und ein F-Test bestätigt die Superiorität der Modellgleichung. Die Simulationsresultate haben großteils Abweichungen von weniger als 5 % bezüglich der Stromstärkemessung. Diese Masterarbeit wurde unterstützt vom Wiener Wissenschafts- und Technologiefonds Projektnummer MA09-028.

Contents

1 Introduction **5**

2 Nanowire Gas Sensors **8**
2.1 Tin Dioxide . . . 10
2.2 Performance Parameters for Gas Sensors . . . 10
2.3 Surface Reaction Mechanisms . . . 11
2.3.1 Response Model for a Mixture of a reducing Gas and Oxygen . . . 12
2.3.2 Adsorption Kinetics . . . 14
2.3.3 Desorption Kinetics . . . 14
2.3.4 Derivation of Model Equations for the Surface Charge Density . . . 15

3 Theory **19**
3.1 Theory of Optimization . . . 19
3.1.1 Optimality Conditions . . . 20
3.2 Optimization Algorithms and their Implementation . . . 22
3.2.1 Line-Search Methods . . . 23
3.2.2 Newton's Method for Optimization . . . 23
3.2.3 Trust-Region Methods . . . 24
3.3 Least-Squares Theory . . . 28
3.3.1 Nonlinear Least-Squares: Estimates and Convergence . . . 28
3.4 Qualitative Analysis of the Model Problem . . . 32
3.4.1 Uniqueness and Existence of the solution . . . 32
3.5 Poisson-Boltzmann Equation . . . 34
3.5.1 Theory of the Poisson-Boltzmann Equation . . . 35
3.5.2 Derivation of the Poisson-Boltzmann Equation from Debye-Hückel Theory . . . 35
3.5.3 Poisson-Boltzmann Equation for Gas Sensor Simulations . . . 42
3.5.4 Graded-channel Approximation . . . 47

4 Simulations **49**
4.1 Numerical Solution of the Surface Equation . . . 49
4.1.1 Results of Parameter Identification . . . 51
4.1.2 Interpretation and Computation of p-Value . . . 56
4.2 Numerical Simulations of the Gas Sensors . . . 58

5 Appendix **63**
5.1 Matlab Code . . . 63

1 Introduction

Gas sensors (gas detectors) are devices that detect the presence of various gases, usually used for safety reasons. In most cases those are battery operated, therefore making them flexible and independent of the power supply. In industry they are either manufactured as portable or stationary (fixed) units and in case of a high level of a certain gas, audible or visible indicators act as an information tool. That means that if a gas passes a certain prespecified level, then an alarm will inform the user about this occurence. While in earlier days, most gas sensors only detected one gas, now they are often multi-functional devices, measuring a wide range of different gases [69].

The market for gas sensors will increase in size over the next decades due to stronger safety regulations implemented worldwide, increasing enhancement of available technology and increasing demand in the chemical industry as well as in the B2C sector, where in particular the medical sector, household sector and fitness and wellness sector (integrated nanosensors in cell phones and sport watches) show high potential according to a market study by WTC (Wicht Technology Consulting) comissioned by AIT (Austrian Institute of Technology), see Figure 1.1, [80].

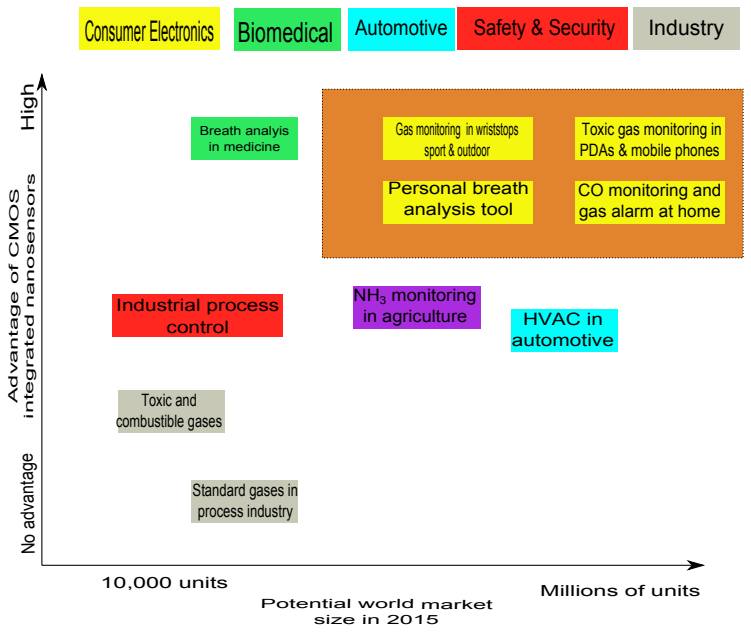

Figure 1.1: Market potential for Nanosensors.

In general, one can distinguish, among others, between the following gas sensors:

- electrochemical sensors,
- metal Oxide Semiconductors,
- catalytic sensors,
- infrared sensors.

Among these are the electrochemical sensor and the metal oxide semiconductor sensors are the most interesting ones.

- **Electrochemical sensors** are commonly used in the detection of toxic gases like CO (carbon monoxide), Cl (chlorine), NO (nitrogen oxide). They work by oxidizing or reducing the target gas at an electrode and measure the resulting current. A further characteristic is their high sensitivity.
- **Metal Oxide Semiconductors**, (**MOS**) are used for the detection of toxic gases as well. (most commonly for CO) and operate via a gas sensitive film, e.g., SnO_2 (tin dioxide). The sensitive film reacts with gases, which causes the device to trigger an alarm if toxic levels are present. In general, MOS are very efficient due to their ability to work in low-humidity ranges. Another advantage is their wide applicability, even being able to detect combustibles.

In this thesis a mathematical model for a nanowire gas sensor will be presented whose gas sensitive film is composed of SnO_2.

Chapter 2 discusses the essential characteristics of the nanowire gas sensors. First the structure of the whole measurement device is described and in the second part the surface reaction equations are derived and presented.

In **Chapter 3** we first present the mathematical theory behind parameter estimation, specifically we discuss general optimization theory and then least squares optimization algorithms that were used for the estimation of the parameter sets. In the second part, we present a uniqueness and existence proof for the surface reaction equations. Furthermore, the nonlinear Poisson-Boltzmann equation is introduced and two different proofs for the uniqueness of its solution are given.

In **Chapter 4** we finally present the simulation results together with a statistical test in which we show that one of the model equations represents the measurement data more accurately than the other equations.

2
Nanowire Gas Sensors

The structure of a metal oxide semiconductor is as given in Figure 2.1 and 2.2. On a silicon substrate (ranging around 400 micrometer) there is a membrane attached, consisting of a membrane layer (of silicon oxide), a heater resistor and thermometer resistor in the middle, and an insulating layer (silicon dioxide (SiO_2) or nitride (N^{3-})) on top, on which the sensing electrodes are positioned. The membrane has a vertical diameter of about 1 to 2 micrometer(μm). Above the sensing electrodes, there is a cover by a metal oxide film, where the surface reaction takes place [50, 17, 6], see Figure 2.3.

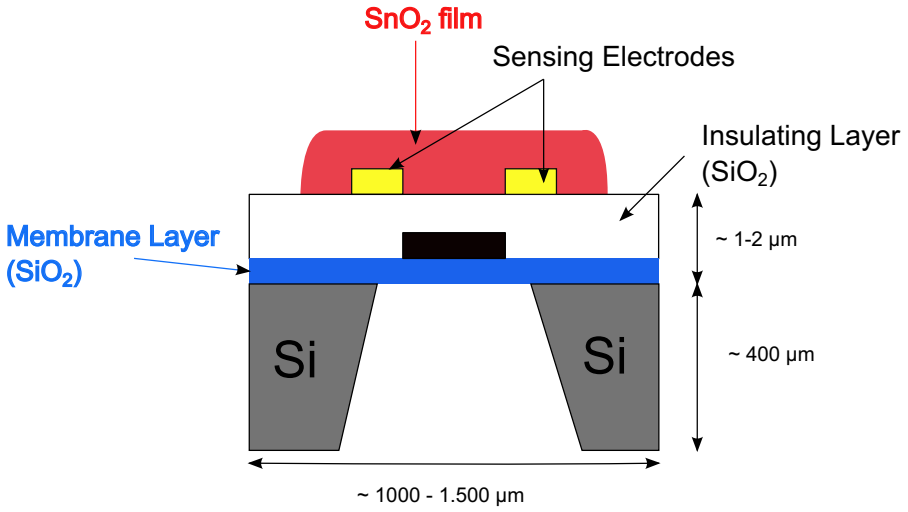

Figure 2.1: Structural design of a modern semiconductor gas sensor.

This sensor chip (diameter around 1 to 1.5 mm) is then put on a device with connections to

the technical device that measures the electric current and is responsible for calling alarm.

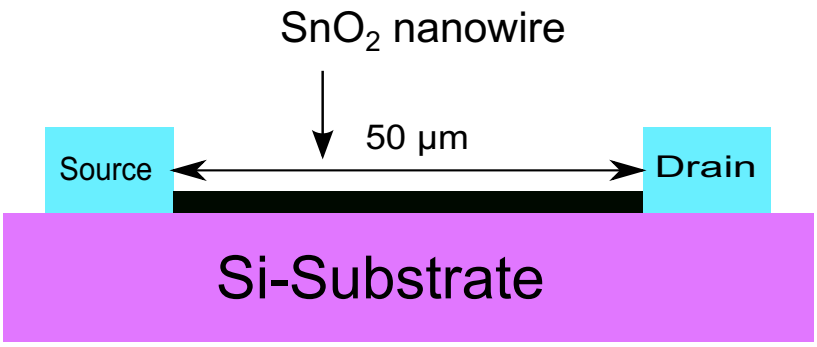

Figure 2.2: Structure of a nanowire sensor.

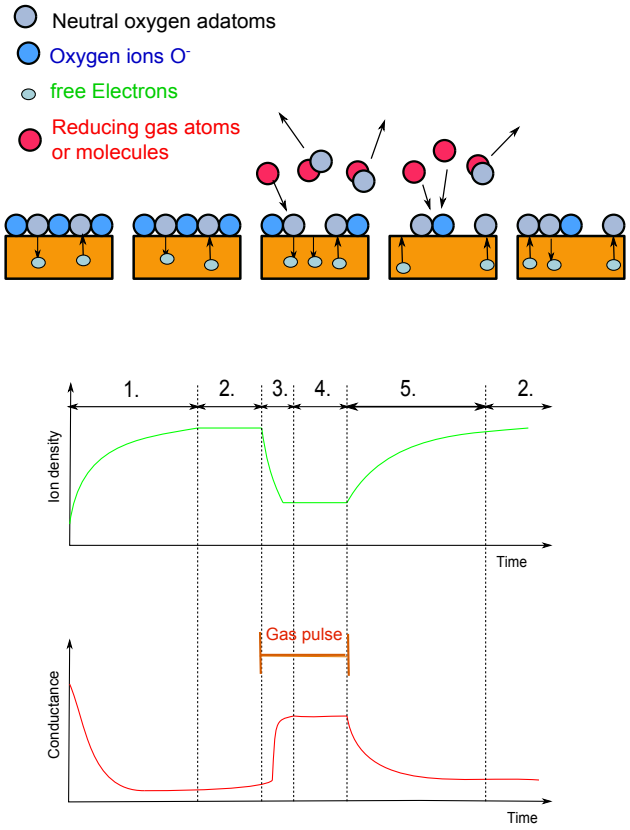

Figure 2.3: Model evolution of a metal-oxide gas sensor. On the top are the physical and chemical processes changing the sensor signal, in the middle is the density of the surface ions, and at the bottom is the sensor signal.

2.1 Tin Dioxide

Tin dioxide (SnO_2) is an inorganic compound, whose mineral form is called cassiterite, the main ore of tin [3]. The colourless, diamagnetic solid occurs naturally by burning tin in air, in particular industrially it is reduced to tin with carbon in a reverbatory furnace burning at 1200 to 1300 degree Celsius [32]. A further important aspect is the fact that it is amphoteric and is as an oxygen-deficient n-type semiconductor [43].

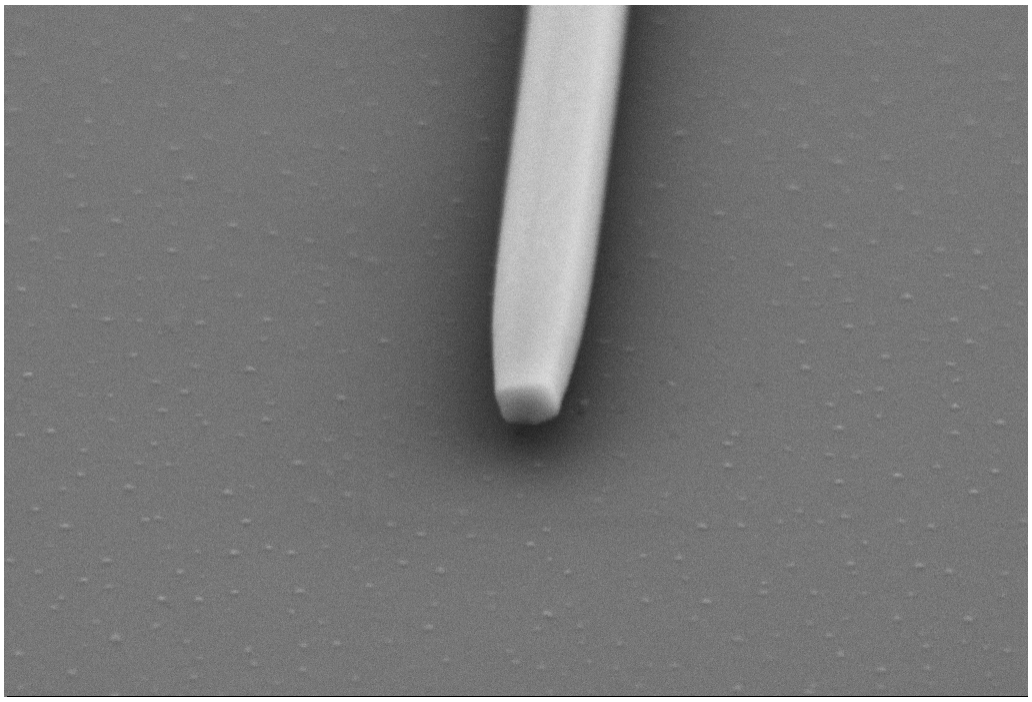

Figure 2.4: Electron-microscope picture of a nanowire produced at the Austrian Institute of Technology. Picture provided by courtesy of the NanoSystems group at the AIT.

2.2 Performance Parameters for Gas Sensors

The purpose of gas sensors is to measure concentrations of one or multiple components of gas mixtures. Therefore, general requirements for the gas sensor are: sensitivity, selectivity, stability, reliability, long lifetime, small size and low cost.

Among others, the literature often refers to the 6S-rule in relation to the desired properties, which are **S**ensitivity, **S**electivity, **S**peed of response, **S**tability, **S**ize/**S**hape [28, 42].

The two most important properties, namely sensitivity and selectivity, are explained in the following.

Sensitivity

The sensitivity S of a sensor is defined as the change of the electric resistance at the time of reaction with the gas compared to the resistance when the sensor is in contact with reference gas

only. For SnO_2 in particular, the sensitivity is

$$S = \frac{R_{ref}}{R_{\text{gas}}} \tag{2.1}$$

for oxidizing gases (e.g., NO_2) and

$$S = \frac{R_{gas}}{R_{ref}} \tag{2.2}$$

for reducing gases (e.g., CO).

In the literature, another definition of sensitivity

$$S := \frac{|R_{gas} - R_{ref}|}{R_{ref}} \tag{2.3}$$

is provided, which returns a better quantification and comparison of the performance of different gas sensors [67], and this definition will be used in the following.

Selectivity The selectivity s_{mn} of a gas sensor is a value that compares the concentration of the distorting substance m with the concentration of the desired analyte n that produces the same sensor signal. This quantification can be achieved by dividing the sensitivity of the sensor with respect to the distorting substance S_m by the sensitivity of the desired substance S_n, i.e., we define the selectivity as

$$s_{mn} := \frac{S_m}{S_n} \tag{2.4}$$

2.3 Surface Reaction Mechanisms

Metal-oxide gas sensors are sensors whose working principle is based on the variation of the conductance due to the presence of oxidizing and reducing gases. The conductivity variations occur due to the interaction between the analyte molecules and the surface of the sensing material [1, 21].

There, adsorbed oxygen ions (O_2^-, O^-, O^{2-}) act as surface acceptors binding electrons from the conduction band of the material. On the other hand, reducing gases (CO) decrease the surface oxygen concentration and result in a decrease of the sensor resistance.

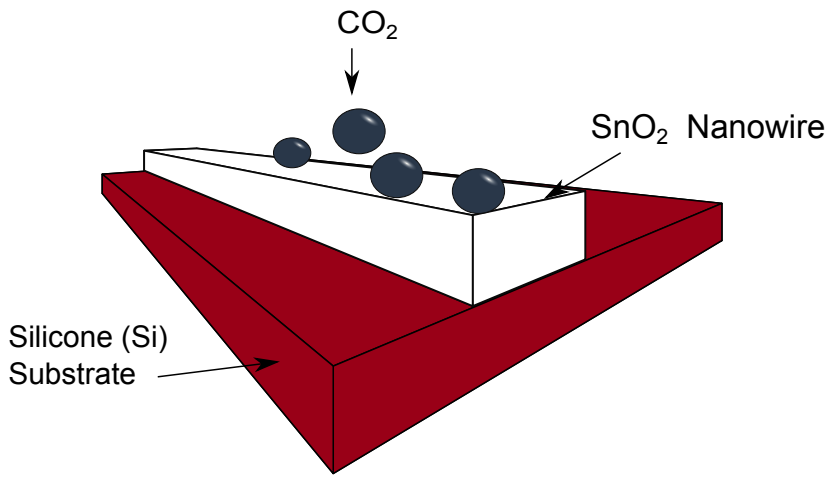

Figure 2.5: A schematic diagram of the nanowire gas sensors considered in the model.

The major problem still being faced today is the low selectivity. A proposed method for improving selectivity is to exploit the dependence of the sensor on the temperature [49, 80].

2.3.1 Response Model for a Mixture of a reducing Gas and Oxygen

Surface interactions as described above are complex and include different reactions [77, 71]. Therefore, in order to arrive at a system of equations that describes the pattern, we need to perform some simplifications. Hence, we will assume that the ionized oxygen O^- is dominant on the tin-oxide surface with respect to other oxygen species [9, 10, 34]. A first approach is to include three chemical processes occuring on the surface. These are

- chemisorption:
$$\frac{1}{2}O_2 + S \rightleftharpoons_{k_{-a}}^{k_a} O - S, \tag{2.5}$$
- ionization:
$$O - S + e^- \rightleftharpoons_{k_{-b}}^{k_b} O^-, \tag{2.6}$$
- and the gas reaction of a reducing gas [77, 10]:
$$O^- - S + R \xrightarrow{k_c} RO + S + e^-. \tag{2.7}$$

We assume that the concentration of the reducing gas $[R]$, e.g. CO (carbon monoxide), is very low compared to the oxygen concentration $[O_2]$ [10, 65].

Notation	Definition
$[S]$	total adsorption site density
$[O_2]$	gas concentration
$N'_S = [O^-]$	ionized oxygen density
$[O-S] = N_O$	neutral adsorbed oxygen density
$[R]$	concentration of the reducing gas
$[RO]$	leaving reducing gas from surface
k_a	adsorption constant for oxygen chemisorption
k_{-a}	desorption constant for oxygen chemisorption
k_b	adsorption constant for oxygen ionization
k_{-b}	desorption constant for oxygen ionization
k_c	adsorption constant reducing gas
n_S	concentration of free electrons

Table 2.1: Notation to express reaction equations (2.9), (2.10), (2.11).

For the transformation of the reaction equations into differential equations we will make use of the Law of mass action [45]:

Law of mass action: *The rate of any given chemical reaction is proportional to the product of the activities (or concentrations) of the reactants.*

In mathematical terms this can be expressed for a general reaction

$$\alpha A + \beta B + \ldots \rightleftharpoons \rho R + \tau T + \ldots \tag{2.8}$$

via

$$\text{forward reaction rate} = k_+[A]^\alpha[B]^\beta \ldots$$
$$\text{backward reaction rate} = k_-[R]^\rho[T]^\tau \ldots$$

where $[A], [B], [R], [T]$ are the active masses and k_+, k_- are the affinity constants [29].

Hence, the three processes can then be expressed by the following differential equations:

$$\frac{dN_O}{dt} = k_a([S] - N_O - N'_S)[O_2]^{1/2} - k_{-a}N_O - \frac{dN'_S}{dt} \tag{2.9}$$

$$\frac{N'_S}{dt} = k_b n_S N_O - k_{-b}N'_S - k_c[R]N'_S \tag{2.10}$$

$$\frac{d[RO]}{dt} = k_c N'_S[R] \tag{2.11}$$

Equation (2.9) describes reaction equation (2.5) that the change in neutral adsorbed oxygen density $\frac{dN_O}{dt}$ equals the adsorption at the unoccupied places $[S] - N_O - N'_S$ minus the desorption at the already occupied places of N_O and minus the change in the ionized oxygen density. Equation (2.10) states in mathematical terms the reaction in (2.6) that the difference in the ionized oxygen

density is equal to the adsorption from the free electrons minus the desorption of the existing occupied surface states and minus adsorption of the reducing gas. Finally, equation (2.11) represents reaction (2.7) where the change in the leaving reducing gas is determined by the adsorption from the surface.

The adsorption and desorption reaction constants play a fundamental role in those equations and are therefore explained in detail in the following.

2.3.2 Adsorption Kinetics

The rate of adsorption R_{ads} of a molecule can be expressed in various ways. One way to express it is via the partial pressure of the molecule in the gas phase above the surface [58]:

$$R_{ads} = k^{ads} P^{\alpha}, \tag{2.12}$$

where k^{ads} is the adsorption rate constant, P is the partial pressure and α is the kinetic order.

If the adsorption rate constant is expressed via the Arrhenius equation, then the k^{ads} can be written in the form

$$k^{ads} = Ae^{-E_a^{ads}/kT} \tag{2.13}$$

where A is the pre-exponential factor, E_a^{ads} is the activation energy for adsorption, k_B is the Boltzmann constant and T is the temperature.

Another way to express the rate of adsorption is given by taking the product of the incident molecular flux F and the sticking probability S_t, meaning that R_{ads} is governed by the rate of arrival of molecules at the surface and the proportion of incident molecules which undergo adsorption [58].

Expressed in mathematical terms, we obtain

$$R_{ads} = S_t F \tag{2.14}$$

where the flux of incident molecules is given by the Hertz-Knudsen equation

$$F = P/(2\pi mkT) \tag{2.15}$$

where $[F] =$ molecules $m^{-2}s^{-1}$, P is the gas pressure, $[P] = Nm^{-2}$, m is the mass of one molecule, $[m] = kg$ and T is the temperature, $[T] = K$.

2.3.3 Desorption Kinetics

Desorption is the process where ions from the surface are returned into the gas phase. In general, the desorbing species is the same as that originally adsorbed assuming that there is no decomposition. However, this is not always the case. For some alkali metals on metallic substrates exhibiting a high work function at low coverages, the desorbing species is the alkali metal ion as opposed to the neutral atom [58].

The rate of desorption R_{des} is given by

$$R_{des} = k^{des} N^{\alpha}, \tag{2.16}$$

where k^{des} is the desorption rate constant for the desorption process, N is the surface concentration of adsorbed species and α is the kinetic order of desorption. The rate constant for the desorption process can again be expressed in Arrhenius form

$$k^{des} = Ae^{-E_a^{des}/kT}, \tag{2.17}$$

where E_a^{des} is the activation energy for desorption and A is the pre-exponential factor which can also be interpreted as the attempt frequency at overcoming the barrier to desorption.

2.3.4 Derivation of Model Equations for the Surface Charge Density

We consider the equations (2.9), (2.10), (2.11) from Section 2.3.1. Those equations are a simplification based on the following assumptions [1]:

1. Only the chemisorbed oxygen is anchored to the surface sites, but the physiosorbed oxygen is not part of the electron exchange process.
2. The chemisorption process is a reaction of first order.

Furthermore, we distinguish between two different assumptions about the ionization reaction: The first assumption (referred to as assumption A) is that the oxygen ionization is the rate-determining step, which implies that the adsorbed neutral oxygen is always at steady state, i.e. $\frac{dN_O}{dt} \approx 0$. The other assumption (referred to as assumption B) is that the chemisorption is a slow process compared to the ionization reaction [76].

Assumption A

Under assumption A we have that $\frac{dN_O}{dt} = 0$ and so the equations simplify to

$$0 = k_a([S] - N_O^{\text{steady}} - N_S')[O_2]^{1/2} - k_{-a}N_O - \frac{N_S'}{dt}, \tag{2.18}$$

$$\frac{N_S'}{dt} = k_b n_S N_O^{\text{steady}} - k_{-b}N_S - k_c[R]N_S', \tag{2.19}$$

$$\frac{d[RO]}{dt} = k_c N_S'[R]. \tag{2.20}$$

If we further assume that the surface coverage is low, i.e., $\frac{[N_O^{\text{steady}} + N_S']}{[S]} \ll 1$, then equation (2.18) becomes

$$0 = k_a[S][O_2]^{1/2} - k_{-a}N_O - \frac{N_S'}{dt}. \tag{2.21}$$

Using equation (2.19) the above equation is equivalent to

$$0 = k_a[S][O_2]^{1/2} - k_{-a}N_O^{\text{steady}} - (k_b n_S N_O^{\text{steady}} - k_{-b}N_S' - k_c[R]N_S)$$

and from this we obtain

$$k_{-a}N_O^{\text{steady}} + k_b n_S N_O^{\text{steady}} = k_a[S][O_2]^{1/2} + k_{-b}N_S' + k_c[R]N_S'$$

and so N_O^{steady} is given by

$$N_O^{\text{steady}} = \frac{k_a[S][O_2]^{1/2} + k_{-b}N'_S + k_c[R]N'_S}{k_{-a} + k_b n_S}. \tag{2.22}$$

We assume that the reaction rate constants are in Arrhenius form [73, 75]

$$k_j = k_{j0} \exp\left(\frac{-E_j}{kT}\right), \tag{2.23}$$

where the parameters are the same as in (2.13).

Via substitution of the Arrhenius form we obtain

$$\frac{dN_S}{dt} = \exp\left(\frac{-N_S^2}{T}\right) \frac{A\exp(-\frac{\omega_1}{T}) + B\exp(-\frac{\omega_2}{T})N_S}{C\exp(-\frac{\omega_3}{T}) + \exp(-\frac{N_S^2}{T})} - B\exp\left(-\frac{\omega_2}{T}\right)N_S - D\exp\left(-\frac{\omega_4}{T}\right)N_S, \tag{2.24}$$

where

$$\begin{aligned}
n_S &= N_D \exp\left(-\frac{q^2 N_S'^2}{2k\epsilon_{dc}\epsilon_0 N_D T}\right), \\
N_S^2 &= \frac{q^2}{2\epsilon_{dc}\epsilon_0 N_D k} N_S'^2, \\
A &= k_{a0}[S][O_2]^{1/2} \frac{q}{\sqrt{2\epsilon_{dc}\epsilon_0 N_D k}}, \\
B = k_{-b0}, \quad C &= \frac{k_{-a0}}{k_{b0}N_D}, \quad D = k_{c0}[R], \\
\omega_1 &= \frac{E_a}{k}, \quad \omega_2 = \frac{E_{-b}}{k}, \\
\omega_3 &= \frac{E_{-a} - E_b}{k}, \quad \omega_4 = \frac{E_c}{k}.
\end{aligned}$$

The concentration of electrons n_S is expressed by

$$n_S = N_D \exp\left(-N_S^2/T\right), \tag{2.25}$$

where we assume that in the considered temperature range all donors are ionized, therefore N_D is constant [34, 9].

$T(K)$	400	500	600	700
$N_D(10^{19})$	1	11	58	260

Table 2.2: Values for N_D as stated in [9].

If we further assume that the variation of ionized oxygen is negligible, then the chemisorbed neutral oxygen can be represented as the equilibrium of the first reaction that is involved in the sensing mechanism, i.e., $\frac{N'_S}{dt} = 0$, provided that we neglect the second reaction effect [1].

Therefore

$$N_O^{\text{steady}} = \frac{k_a[S][O_2]^{1/2}}{k_{-a}} = \frac{k_{a0}[S][O_2]^{1/2}}{k_{-a0}} \exp\left(-\frac{E_a - E_{-a}}{kT}\right) \tag{2.26}$$

and so we obtain the equation

$$\frac{dN_S}{dt} = A \exp\left(-\frac{N_S^2 + \omega_1}{T}\right) - B \exp\left(-\frac{\omega_2}{T}\right) N_S - D \exp\left(-\frac{\omega_4}{T}\right) N_S, \tag{2.27}$$

where

$$A = \frac{k_{a0}}{k_{-a0}}[S] N_D k_{b0} \frac{q}{\sqrt{2\epsilon_{dc}\epsilon_0 N_D k}},$$
$$\omega_1 = \frac{E_a - E_{-a} + E_b}{k}$$

and the other terms are as in equation (2.25).

Assumption B

If we suppose that the chemisorption is a slow process when compared to the ionization reaction, then we make the following simplification [1]: We assume that the adsorbed oxygen concentration is small in the interesting temperature range and observation time and modify equation (2.9) to

$$N_O = N_{ox} - N_S', \tag{2.28}$$

where N_{ox} is the concentration of all the oxygen species on the surface, and is considered constant in the model. So the whole system becomes

$$N_O = N_{ox} - N_S', \tag{2.29a}$$
$$\frac{N_S}{dt} = k_b n_S N_O - k_{-b} N_S' - k_c[R] N_S', \tag{2.29b}$$
$$\frac{d[RO]}{dt} = k_c N_S'[R] \tag{2.29c}$$

and a similar derivation to the one in hypothesis A now leads to

$$\frac{N_S}{dt} = A \exp\left(-\frac{N_S^2 + \omega_1}{T}\right) - B \exp\left(-\frac{N_S^2 + \omega_1}{T}\right) N_S - C \exp\left(-\frac{\omega_2}{T}\right) N_S - D \exp\left(-\frac{\omega_4}{T}\right) N_S, \tag{2.30}$$

where

$$A = N_{ox} N_D k_{b0} \frac{q}{\sqrt{2\epsilon_{dc}\epsilon_0 N_D k}},$$
$$\omega_1 = \frac{E_b}{k},$$
$$B = N_D k_{b0}$$

and the other parameters as in equation (2.25) [1].

For the models we now must estimate the parameters $A, B, C, D, \omega_1, \omega_2, \omega_3, \omega_4$ from the data.

Parameter	A	B	C	D	ω_1	ω_2	ω_3	ω_4
Unit	$\mathrm{K}^{1/2}\mathrm{s}^{-1}$	s^{-1}	s^{-1}	s^{-1}	K	K	K	K

Table 2.3: Estimation parameters and their corresponding units (K = Kelvin, s = seconds) [2].

We again consider the equation (2.25) that represents the relationship between the concentration of free electrons n_S and the occupied surface states N_S. Due to the relation $n_S = \frac{Is}{q}$ where I is the current, from the measurements $I = 2.0 \times 10^{-6}$ (s is seconds, assumed to be 1) and $q = 1.60217646 \times 10^{-19}$, the elementary charge, we are able to transform equation (2.25) into

$$N_S = \sqrt{-T \log\left(\frac{n_S}{N_D}\right)}. \tag{2.31}$$

From this representation and the measurement values we are now able to obtain values for the occupied surface states N_S where a selection of values are represented in Table 2.4.

Time (s)	N_S	**Temperature** (K)
1	84.720	453.15
1544	73.068	372.15
3396	73.068	373.15
5610	77.272	423.15
8171	85.083	462.15
11080	85.682	473.15
14310	92.618	523.15
17880	99.495	573.15
21832	105.990	623.15
26125	112.570	673.15

Table 2.4: Time and corresponding N_S value for data provided by the AIT.

We will add in the equations (2.27), (2.30) and (2.24) another parameter K that accounts for different reaction behaviors. Due to the complex interactions and sometimes random behavior of chemical reactions, we might could also assume K to be a random variable, but this would go beyond the focus of this thesis and is up to further research.

3 Theory

3.1 Theory of Optimization

Optimization problems can be expressed as finding

$$\begin{aligned} &\min_x f(x) \\ &s.t.\, x \in C \subset \mathbb{R}^n \end{aligned}$$

where f is the function to minimize, called *objective function*, and C is the set of constraints.

A *local minimizer* x^* is an element $x \in C$ with

$$f(x^*) \leq f(x) \tag{3.1}$$

for all $x \in C \cap B_\delta(x^*)$ for a $\delta > 0$. x^* is a *strict local minimizer* if there exists $\delta > 0$, such that

$$f(x^*) < f(x) \tag{3.2}$$

for all $x \in C \cap B_\delta(x^*)$. Furthermore, if

$$f(x^*) \leq f(x) \tag{3.3}$$

for all $x \in C$, then we call x^* a *global minimum.*

We will use the following notation: $e = x^* - x$ is the error, $e_n = x^* - x_n$ the error of the n-th iterate, and $\mathcal{B}(r)$ the ball of radius r around x^*

$$\mathcal{B}(r) := \{x | \, ||e|| < r\}.$$

The *gradient* of f is defined by $\nabla f(x) := (\partial f/\partial x_1, \ldots, \partial f/\partial_n)$ and the *Hessian matrix* of f by $H_f(x) := (\partial^2 f/\partial x_i \partial x_j)$. We will assume that f is twice continuously differentiable, implying symmetry of the Hessian, i.e., $H_f = H_f^T$.

The normal Euclidean norm is denoted by $||x|| := \sqrt{\sum_{i=1}^{d} x_i^2}$ and the matrix norm is given by

$$||A|| := \sup_{x \neq 0} \frac{||Ax||}{||x||}.$$

Definiteness of the matrix defined here is important for stating the optimality conditions of a problem.

Definition 3.1.1. A $d \times d$ matrix A is positive semidefinite (psd) if $x^T Ax \geq 0$ for all $x \in \mathbb{R}^d$ and if there exists $x \neq 0$ such that $x^T Ax = 0$. A is called positive definite (pos def) if $x^T Ax > 0$ for all $x \neq 0$. If there exist $x, y \in \mathbb{R}^d$ with $x^T Ax > 0$ and $y^T Ay < 0$, then A is called indefinite.

Another important property of a function is convexity:

Definition 3.1.2. A function $f : \mathbb{R}^d \to \mathbb{R}$ is called *convex* if for all $x, y \in \mathbb{R}^d$ and $\lambda \in [0, 1]$ the inequality

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda) f(y) \tag{3.4}$$

holds.

Every linear function is convex and functions of the form $f(x) = x^2$ are convex.

3.1.1 Optimality Conditions

Before giving a necessary condition for a minimizer, we prove a theorem that defines the notion of a *descent direction* being later of prime importance for iterative methods as well as optimality statements.

Theorem 3.1.3. *Let $f : \mathbb{R}^d \to \mathbb{R}$ be differentiable at the point $\hat{x}$. If there exists a vector $p \in \mathbb{R}^d$ such that $\nabla f(\hat{x})^T p < 0$, then there exists $\epsilon > 0$ such that $f(\hat{x} + \epsilon p) < f(\hat{x})$ for all $\theta \in (0, \epsilon)$. The vector p is called a descent direction [12].*

Proof. From differentiability of f, we have

$$f(\hat{x} + \theta p) = f(\hat{x}) + \theta \nabla f(\hat{x})^T p + o(\theta).$$

Rearranging the terms and division by θ ($\theta \neq 0$), we obtain

$$\frac{f(\hat{x} + \theta p) - f(\hat{x})}{\theta} = \nabla f(\hat{x})^T p + o(1).$$

As $\nabla f(\hat{x})^T p < 0$ and $o(1) \to 0$ for $\theta \to 0$, there exists $\epsilon > 0$ such that

$$\nabla f(\hat{x})^T p + o(1) < 0$$

for all $\theta \in (0, \epsilon)$ and so

$$\frac{f(\hat{x} + \theta p) - f(\hat{x})}{\theta} < 0$$

implying $f(\hat{x} + \theta p) - f(\hat{x}) < 0$. □

Remark: If $\nabla f(x) \neq 0$ then we always obtain a descent direction by setting $p := -\nabla f(x)$, because $\nabla f(x)p = -\nabla f(x)^T \nabla f(x) = -\,||\nabla f(x)|| < 0$.

Next we will prove a necessary condition for an optimal point without constraints.

Theorem 3.1.4. ***Necessary optimality condition for unconstrained optimization.*** *Let $f : \mathbb{R}^d \to \mathbb{R}$ be twice differentiable and x^* a local minimizer of f. Then*

$$\nabla f(x^*) = 0. \tag{3.5}$$

Moreover, $H_f(x^*)$ is psd [40].

Proof. Let $v \in \mathbb{R}^d$, then we obtain from Taylor's theorem that for sufficiently small $s > 0$

$$f(x^* + sv) = f(x^*) + s\nabla f(x^*)^T v + \frac{s^2}{2} v^T H_f(x^*) v + o(s^2).$$

As x^* is a local minimizer we must have for sufficiently small s that

$$0 \geq f(x^* + sv) - f(x^*)$$

and hence

$$\nabla f(x^*)^T v + \frac{s}{2} v^T H_f(x^*) v + o(s) \geq 0$$

for all sufficiently small $s > 0$ and $v \in \mathbb{R}^d$. Letting $s \to 0$ and $v = -\nabla f(x^*)$ we obtain

$$||\nabla f(x^*)||^2 = 0$$

Setting $\nabla f(x^*) = 0$, dividing by s, and letting $s \to 0$, we find

$$\frac{1}{2} v^T H_f(x^*) v \geq 0$$

for all $v \in \mathbb{R}^d$. This concludes the proof. □

The condition $\nabla f(x) = 0$ is called a *first-order optimality condition* and the condition H_f psd is called a *second-order optimality condition.*

For general differentiable functions there does not exist a first-order sufficient optimality condition. (Consider the function $f(x) = x^3$; at $x = 0$ we have $\nabla f(0) = 0$ but the function does obviously not have a local minimum there.) So we need to resort to a second-order condition.

Theorem 3.1.5. *Suppose $f : \mathbb{R}^d \to \mathbb{R}$ is twice differentiable at x^*. If $\nabla f(x^*) = 0$ and $H_f(x^*)$ is positive definite, then x^* is a strict local minimum.*

Proof. First we have, for all $x^* \in \mathbb{R}^d$,

$$f(x) = f(x^*) + \nabla f(x^*)^T (x - x^*) + \frac{1}{2}(x - x^*)^T H_f(x^*)(x - x^*) + o(||x - x^*||^2), \tag{3.6}$$

where again $o(||x - x^*||^2) \to 0$ for $x \to x^*$. We will prove the statement by contradiction. Assume that x^* is not a strict local minimum. Then there exists a sequence $\{x_k\}$ converging to x^* such that $f(x_k) \leq f(x^*)$, $x_k \neq x^*$, for each k.

Set $p_k := \frac{x_k - x^*}{||x_k - x^*||}$, then we obtain from equation (3.6) that $\nabla f(x^*) = 0$ and from $f(x_k) \leq f(x^*)$ that

$$\frac{1}{2}(p_k)^T H_f(x^*)(p_k) + o(1) \leq 0.$$

From $||p_k|| = 1$ for all k, we obtain an index set $J \subset \mathbb{N}$ such that $\{p_k\}_J$ converges to p where $||p|| = 1$. Considering the subsequence and the fact that $o(1) \to 0$ for $k \to \infty$, we obtain $p^T H_f p \leq 0$. This implies a contradiction to our assumption that $H_f(x^*)$ is positive definite since $||p|| = 1$. Therefore, x^* is a strict local minimum.

As f is twice continuously differentiable, we obtain that $H_f(x)$ is positive definite in a ball $\mathcal{B}(\epsilon)$ for sufficiently small ϵ, and so f is strictly convex in an ϵ-ball around x^*. Therefore, x^* is the global minimum over $\mathcal{B}_\epsilon(x^*)$ because for any $x \in \mathcal{B}_\epsilon(x^*)$ holds $f(x) = f(x^*) + \frac{1}{2}(x - x^*)^T H_f(x^*)(x - x^*) + o(||x - x^*||^2) > f(x^*)$ due to the positive definite of the H_f, which proves the statement. [12] □

If we know that the function to minimize is convex, then we obtain the first-order sufficient optimality condition stated below.

Theorem 3.1.6. *Let $f : \mathbb{R}^d \to \mathbb{R}$ be convex and continuously differentiable. Then is x^* a global minimum if and only if $\nabla f(x^*) = 0$.*

Proof. If x^* is a global optimum, then it follows from Theorem 3.1.4 that $\nabla f(x^*) = 0$. Now assume that $\nabla f(x^*) = 0$. We first show that for convex differentiable f the inequality

$$f(x) \geq f(y) + \nabla f(y)(x - y)$$

holds for all $x, y \in \mathbb{R}^d$. From convexity of f, it first follows that

$$f(sx + (1 - s)y) \leq sf(x) + (1 - s)f(y)$$

and via rearrangement we obtain

$$f(y) + \frac{f(sx + (1 - s)y) - f(y)}{s} \leq f(x)$$

for $s \in (0, 1)$. As this holds for all $s \in (0, 1)$ so follows that for $s \to 0$ and the Taylor expansion of $f(sx + (1 - s)y)$ we have

$$f(y) + \nabla f(y)(x - y) \leq f(x)$$

which shows the intermediate result. But if $\nabla f(x^*) = 0$, then

$$f(x) \geq f(x^*) + \nabla f(x^*)(x - x^*) = f(x^*)$$

for all $x \in \mathbb{R}^d$ and so x^* is a global minimum. □

3.2 Optimization Algorithms and their Implementation

The results and theorems provided in Section 3.1 are important for the qualitative and theoretical analysis of optimization problems. However, finding points that satisfy the corresponding conditions is a different business.

In particular, almost all optimization algorithms will be iterative and so one solely approximates the solution of an optimization problem. Therefore starting point, stability, direction of movement and Jacobian and Hessian information are of utmost importance.

The idea of iterative algorithms is to choose a starting point x_0 and an initial direction p_0 in order to get to a new point x_1. One then repeats the procedure until the decrease in the function value becomes smaller than a certain threshold.

Nevertheless, the concept of *descent direction* will again be important and while it is reasonable to think that if one always goes in the opposite direction of the gradient of a function – this method is called *steepest descent* – it will be shown that in practice it is quite often a very poor choice, i.e., it converges slowly.

The goal of this chapter is to introduce first the line-search and further the trust-region method, which are vital for the optimization problem we consider in Section 3.3.

3.2.1 Line-Search Methods

The general idea behind line-search methods is to minimize for a given point x_k and a direction p_k the function

$$f(x_k + \alpha p_k) \tag{3.7}$$

with respect to α.

Writing $I(\alpha) := f(x_k + \alpha\, p_k)$, we have a necessary condition that $I'(\alpha) = 0$. As $I'(\alpha) = p_k^T \nabla f(x_k + \alpha p_k)$, this implies that we have to solve $p_k^T \nabla f(x_k + \alpha p_k) = 0$ [12].

Definition 3.2.1. A line-search method is called *exact* if the parameter is chosen such that the function $I(\alpha) := f(x_k + \alpha p_k)$ takes on a global minimum. Otherwise, we call the method *inexact*.

Generally, solving the last equation resorts to solving for the zeros of a nonlinear function, being in most cases computationally very expensive. Moreover, in most cases the additional gain of finding the global minimizing parameter α and a sufficient decrease parameter β is quite small, and so one often resorts to an inexact line-search, among which the *Armijo rule* is the most significant [12, 40].

Definition 3.2.2. The *Armijo rule* is to find a step size parameter α such that

$$f(x_k + \alpha p_k) < f(x_k) + \theta \alpha \nabla f(x_k)^T p_k \tag{3.8}$$

where $\theta \in (0, 1)$ is an algorithmic parameter [12]. For example J. Nocedal [59] suggests $\theta = 10^{-4}$.

3.2.2 Newton's Method for Optimization

Newton's method for optimization is an algorithm to find stationary points of function. It essentially uses Hessian information for choosing its step sizes. For a comparison of Newton and Gradient descent see Figure 3.1.

The idea behind it is the following: The second order Taylor expansion of a function f is given by

$$f(x_k + p_k) = f(x_k) + \nabla f(x_k) p_k + \frac{1}{2} p_k^T H_f(x_k) p_k,$$

where we set $x_{k+1} - x_k =: p_k$. Differentiating the function with respect to p_k and setting $\nabla f(x_k + p_k) = 0$, we obtain

$$\nabla f(x_k) + H_f(x_k)p_k = 0$$

and furthermore

$$x_{k+1} = x_k - [H_f(x_k)]^{-1}\nabla f(x_k) \tag{3.9}$$

assuming that $H_f(x_k)$ is invertible [54].

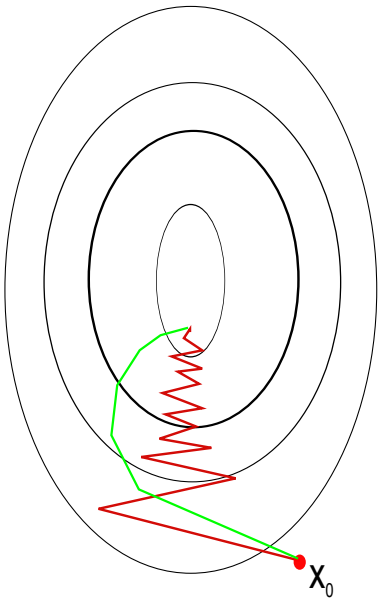

Figure 3.1: A comparison the Newton method (in green) and Gradient descent (in red) for minimizing a function. It can be seen that the Newton method essentially uses curvature information in order to obtain a more direct route.

Newton methods have been extended to quasi-Newton methods where the Hessian is approximated via gradient information, because computing the Hessian is prohibitively expensive for large optimization methods.

3.2.3 Trust-Region Methods

Despite the ease with which a line-search method is implemented, in general line-searches perform quite poorly for non-convex functions resp. functions with non-semi positive definite approximate Hessians [12]. Therefore, *trust region* methods overcome the problems that line-search methods encounter with non-semi positive definite approximate Hessians.

The idea behind trust-region methods is the following: Let r be the radius of the ball around x_c where the quadratic model

$$m_{qc}(x) = f(x_c) + \nabla f(x_c)^T(x - x_c) + \frac{1}{2}(x - x_c)^T H_c(x - x_c) \tag{3.10}$$

accurately represents the function f in the ball with radius r. We call r the *trust-region radius* and the ball

$$\mathcal{B}(r) = \{x| \, ||x - x_c|| \leq r\} \tag{3.11}$$

is called the *trust region*.

We then proceed by minimizing m_{qc} over $\mathcal{B}(r)$ in order to obtain a point x_+, in mathematical terms we try to find

$$\min_{\tau \in \mathcal{B}(r)} m_{qc}(\tau). \tag{3.12}$$

Setting $s := x_+ - x_c$, we then need to decide whether to accept the step and/or to change the trust region radius. Intuitively, we would accept a step when $f(x_+) = f(x_c + s) < f(x_c)$ but we will see that this condition can be generalized and made more flexible.

We will now state some conditions when to accept a step and when not. Define the *actual reduction* of f by

$$R_{act} = f(x_c) - f(x_+) \tag{3.13}$$

and the *predicted reduction*

$$R_{pred} = m_{qc}(x_c) - m_{qc}(x_+) = -\nabla f(x_c)^T s - \frac{1}{2} s^T H_c s. \tag{3.14}$$

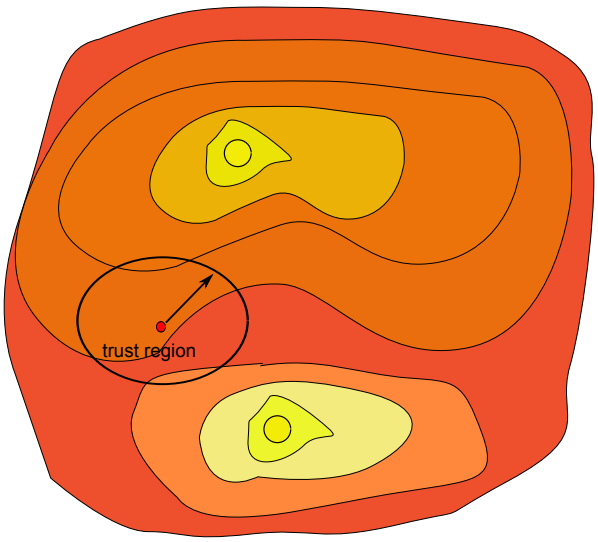

Figure 3.2: Trust region.

We will further use the following three control parameters: λ_0, λ_{low}, λ_{high} where

$$\lambda_0 \leq \lambda_{low} < \lambda_{high}. \tag{3.15}$$

Now we consider the following cases [40]:

1. If $R_{act}/R_{pred} < \lambda_0$, then the step should be rejected.

2. If $R_{act}/R_{pred} < \lambda_{low}$, then the trust region radius should be decreased.

3. If $\lambda_{low} \leq R_{act}/pred \leq \lambda_{high}$, then accept the step and keep the current trust region radius.

4. If $R_{act}/R_{pred} > \lambda_{high}$, then the trust region radius should be increased.

The advantage of this approach over the former, where we would have accepted a step if

$$f(x_+) < f(x_c),$$

is that the quality of the quadratic approximation is also taken into account.

The contraction and expansion of the trust region radius are then obtained by multiplying r by constants

$$0 < \omega_{down} < 1 < \omega_{up} \tag{3.16}$$

where typically $\omega_{down} = 1/2$ and $\omega_{up} = 2$.

Nonlinear least squares problem usually make use of one of the following two algorithms:

- Trust-Region-Reflective,
- Levenberg-Marquardt method.

Trust-Region-Reflective Trust-Region-Reflective, also called subspace trust region method, is a trust-region method that reduces the search space of the trust-region subproblem

$$\min_{||Ds||_2 \leq r} (g_c^T s + \frac{1}{2} s^T H_c s) \tag{3.17}$$

where D is a diagonal scaling matrix, g_c is the gradient of f at x_c, H_c the Hessian of f at x_c and r is the radius of the ball to a two-dimensional subspace S [36].

This is done because in general algorithms solving (3.17) involve the computation of a full eigensystem and applying a Newton method to the equation

$$\frac{1}{r} - \frac{1}{||s||_2} = 0. \tag{3.18}$$

The major problem of these algorithms is that they require time proportional to a large number of factorizations of the Hessian H_c. Therefore, if the problem is higher dimensional, an approximation is needed, which is in our case a two-dimensional subspace [51, 63]. In particular, if the two-dimensional subspace S is already computed then solving problem (3.17) is trivial. So the major work now goes into the appropriate determination of the subspace S. For the trust-region subproblem we define the two-dimensional subspace S as the linear space spanned by s_1 and s_2, where s_1 is the direction of the gradient g_c and s_2 is either

$$H_c s_2 = -g_c, \tag{3.19}$$

or a direction with

$$s_2^T H_c s_2 < 0. \tag{3.20}$$

For the specific case of data fitting we consider the problem

$$\min_{x\in\mathbb{R}^n} \sum_i f_i(x)^2 = \min_{x\in\mathbb{R}^n} ||F(x)||_2^2 \tag{3.21}$$

with

$$F(x) = \begin{pmatrix} f_1(x) \\ \vdots \\ f_m(x) \end{pmatrix}$$

where m corresponds to the number of data points. In order to determine the two-dimensional subspace S, we will make use of an approximate Gauss-Newton direction s given as the solution of

$$\min_s ||Js + F||_2^2 .$$

Levenberg-Marquardt Algorithm (LMA) Consider the linear approximation of the function (3.22) given by

$$F(x_k + p_k) \approx F(x_k) + J(x_k)p_k.$$

We want to minimize

$$||J(x_k)p_k + F(x_k)||_2^2$$

with respect to p_k, therefore the minimizer is given by the solution of the normal equation

$$J(x_k)^T J(x_k)p_k = -J(x_k)^T F(x_k). \tag{3.22}$$

Kenneth Levenberg [44] replaced equation (3.22) by a damped version

$$(J(x_k)^T J(x_k) + \lambda I)p_k = -J(x_k)^T F(x_k), \tag{3.23}$$

where I is the identity matrix and λ is a parameter that can be adjusted at each iteration such that if the descent is rapid, then λ is chosen smaller and if we have insufficient descent then λ is larger. That means that for small λ, the direction is close to Gauss-Newton direction, while if λ is large then it is closer to the gradient descent direction.

Marquardt in 1963 realized that by scaling each component of the gradient according to the curvature, there will be a larger movement along directions with small gradient [52]. The Levenberg-Marquardt direction p_k is then given as the solution of the equation

$$\left(J(x_k)^T J(x_k) + \lambda \operatorname{diag}(J(x_k)^T J(x_k))\right) p_k = -J(x_k)^T F(x_k) \tag{3.24}$$

where again λ is chosen according to whether a close Gauss-Newton direction or a scaled version of the gradient descent direction is desired [27].

3.3 Least-Squares Theory

Least squares is a standard approach in finding an approximation to an overdetermined system of equations. It satisfies the condition that the sum of the square residuals becomes minimal. A standard example of using least squares is the case of fitting a line to a given set of points, called linear regression. This method still prevails in many application areas of business, economics, biology, chemistry and engineering, where one wants to extract some information between the variables on the x-axis and the y-axis [79].

In general, one can distinguish between linear least squares and non-linear least squares, depending on whether the residuals are all linear in all unknowns or not.

- **Linear least squares** implies that the residuals are all linear in the unknowns. There is a closed form for the solution and it occurs mostly in statistical regression analysis, meaning that one has a set of data points and fits a function (usually a polynomial of degree 1, 2 or 3 where the optimization is performed over the set of all coefficients) to best approximate the given data set. In particular this means that the problem can be reformulated into solving

$$Au = b, \tag{3.25}$$

  where u is the sought solution vector, A is the corresponding coefficient matrix, and b is a vector of values.

- **Nonlinear least squares** implies that the residuals are not all linear in the unknowns. In general, one does not have a closed form for the solution and needs to apply an iterative procedure in order to obtain a solution. Furthermore, there might exist several minimums and to find the global minimum might require high computational costs. In general, the problem is only given in the form $F(u) = y$ where F is the defining function and y is the vector of values [72].

For the case of estimating parameters for the ODE's governing the development of the density of electrons, we are mostly concerned with non-linear problems and therefore will restrict the discussion to non-linear least squares. In fact, all methods applied to the non-linear case are applicable to the linear case as well, except that the direct methods applicable to linear problems are in general much faster.

3.3.1 Nonlinear Least-Squares: Estimates and Convergence

Assume there is a given data set (x_i, y_i) for $i = 1, \ldots, n$ with a known functional relationship f where

$$y_i = f(x_i, \theta) + r_i(\theta), \tag{3.26}$$

x_i is a $k \times 1$ vector, θ belongs the parameter set $\Theta \subset \mathbb{R}^d$, and the function f depends nonlinearly on θ. Then *nonlinear least squares* problems are of the form

$$f(x) = \frac{1}{2}\sum_{i=1}^{n} ||r_i||_2^2 = \frac{1}{2}R(x)^T R(x) \tag{3.27}$$

where the vector $R = (r_1, \ldots, r_n)$ is the *residual vector*.

If d is the number of parameters, then for the case $n > d$ the problem is *overdetermined*, if $d = n$ we have a system of equations to solve (therefore standard numerical methods for the solution of a system of equations can be applied, e.g., see [39]) and if $d > n$ the problem is *underdetermined*. The case of underdetermination is rare, nevertheless it is important for the solution of high-index differential algebraic equations [68, 15].

If x^* is a local minimizer of f and $f(x^*) = 0$, then the problem of minimizing f is called a *zero-residual problem*. If $f(x^*)$ is small, then the problem is called *small-residual problem*, and otherwise it is called a *large-residual problem*.

The Jacobian of R is given by J and so

$$\nabla f(x) = J^T(x)R(x) \in \mathbb{R}^d. \tag{3.28}$$

From the optimality condition in Theorem 3.1.4 we obtain that for a minimizer x^* the equation

$$J(x^*)^T R(x^*) = 0 \tag{3.29}$$

holds. For the underdetermined case and if $J(x^*)$ has full row rank, then $R(x^*)$ must be zero.

This condition does not necessarily hold for the overdetermined case, therefore leaving a much greater choice of values.

The Hessian of f can then be expressed via

$$H_f(x) = J(x)^T J(x) + \sum_{i=1}^{n} r_i^T H_{r_i}(x) \tag{3.30}$$

which in particular requires the computation of n Hessians H_{r_i}, which is in practice too costly and therefore needs to be approximated. Furthermore, we have

$$\sum_{i=1}^{n} r_i(x)^T H_{r_i}(x) = R''(x)^T R(x)$$

where the second derivative R'' of R is a tensor.

Gauss-Newton Method

For least squares problem, one is in general able to exploit the structure of the problem in order to develop a more efficient algorithm for the solution of the minimum of the sum of squares. Gauss-Newton is an algorithm that is focused on the minimization of the sum of the squares of functions without requiring second derivatives, which are in many applications very expensive to compute [25, 13].

The algorithm works as follows:

1. Choose an appropriate initial value x_0 for the minimizer.
2. Iterate until $||p|| < \delta_{tol}$
   (a) Compute the Jacobian J_k at x_k.

(b) Solve the *normal equations*

$$(J_k^T J_k)p = J_k^T r, \tag{3.31}$$

where r is the vector of the residual functions r_i.

(c) Update $x_{k+1} = x_k + p$.

We will now state an error estimate for overdetermined problems.

Theorem 3.3.1. *Let $n > d$ and $J(x^*)$ have full column rank. Then there exist $C > 0$ and $\epsilon > 0$ such that for $\mathcal{B}_\epsilon(x^*)$ the error in the Gauss-Newton iteration satisfies*

$$||e_+|| \leq C(||e_c||^2 + ||J(x^*)||\,||e_c||). \tag{3.32}$$

Proof. [40] Let $\epsilon > 0$ such that $||x - x^*|| < \epsilon$ implies $J(x)^T J(x)$ is non-singular. Further let γ be the Lipschitz constant for J. From (3.31) we obtain

$$\begin{aligned} e_+ &= e_c - (J(x_c)^T J(x_c))^{-1} J(x_c)^T R(x_c) \\ &= (J(x_c)^T J(x_c))^{-1} J(x_c)^T (J(x_c)e_c - R(x_c)). \end{aligned}$$

Furthermore, we obtain

$$\begin{aligned} J(x_c)e_c - R(x_c) &= J(x_c)e_c - R(x^*) + R(x^*) - R(x_c) \\ &= -R(x^*) + (J(x_c)e_c + R(x^*) - R(x_c)). \end{aligned}$$

In particular,

$$||J(x_c)e_c + R(x^*) - R(x_c)|| \leq \gamma\,||e_c||^2\,/2$$

and from the optimality condition (3.29), i.e., $J(x^*)^T R(x^*) = 0$, we get

$$-J(x_c)^T R(x^*) = (J(x^*) - J(x_c))^T R(x^*).$$

Therefore,

$$\begin{aligned} ||e_+|| &\leq \left|\left|(J(x_c)^T J(x_c))^{-1}\right|\right| \left|\left|(J(x^*) - J(x_c))^T R(x^*)\right|\right| \\ &\quad + \frac{\left|\left|(J(x_c)^T J(x_c))^{-1}\right|\right| \left|\left|J(x_c)^T\right|\right| \gamma\,||e_c||^2}{2} \\ &\leq \left|\left|(J(x_c)^T J(x_c))^{-1}\right|\right| \gamma\,||e_c|| \left(\frac{||R(x^*)|| + \left|\left|J(x_c)^T\right|\right| ||e_c||}{2}\right). \end{aligned}$$

Setting

$$C := \gamma \max_{x \in \mathcal{B}_\epsilon(x^*)} \left|\left|(J(x)^T J(x))^{-1}\right|\right| \left(\frac{1 + \left|\left|J(x)^T\right|\right|}{2}\right) \tag{3.33}$$

we obtain (3.32). □

A consequence from Theorem 3.3.1 is that for zero-residual problems, the term $||R(x^*)||\,||e_c||$ in (3.32) vanishes and therefore the local convergence rate is q-quadratic. Furthermore for $x_c \in \mathcal{B}_\epsilon(x^*)$ we obtain $||e_+|| \leq \delta\,||e_c||$ for some $\delta \in (0,1)$ if

$$C(\epsilon + ||J(x^*)||) \leq \delta$$

and therefore the q-factor equals $C\,||J(x^*)||$. Although we cannot even guarantee q-linear convergence for small and large residuals, Gauss-Newton will be fast for small residuals and good initial data. For large residual problems however, Gauss-Newton unfortunately may not converge at all.

Consider again the overdetermined least-squares objective function

$$f(x) = \frac{1}{2}\sum_{i=1}^{n} ||r_i(x)||_2^2 = \frac{1}{2}R(x)^T R(x).$$

Then the direction of steepest descent is the negative of the gradient, i.e.,

$$-\nabla f(x) = -J(x)^T R(x) \tag{3.34}$$

because the gradient of a function is the direction of steepest increase.

The Gauss-Newton direction at x is then given as in (3.31) by

$$p = -(J(x)^T J(x))^{-1} J(x)^T R(x) \tag{3.35}$$

and for J having full column rank, we obtain

$$p^T \nabla f(x) = -(J(x)^T R(x))^T (J(x)^T J(x))^{-1} J(x)^T R(x) < 0$$

which in particular implies that the Gauss-Newton direction is a descent direction. If we now choose the step size parameter α such that

$$f(x + \alpha p) \leq f(x) + c\alpha p^T \nabla f(x) \tag{3.36}$$

and update x by αp, then we obtain the *damped Gauss-Newton* iteration [59]. In particular, damped Gauss-Newton is Gauss-Newton together with the Armijo rule.

In general, damped Gauss-Newton is an effective algorithm when the matrices $(J(x_k)^T J(x_k))$ have full column rank and are uniformly bounded and well conditioned, which is a very strong assumption.

In order to make the algorithm applicable more widely, one adds a regularization parameter $\lambda > 0$ to $J(x_c)^T J(x_c)$ to obtain $x_+ = x_c + p$, where

$$p = -(\lambda_c I + J(x_c)^T J(x_c))^{-1} J(x_c)^T R(x_c),$$

I is the identity matrix, and $\lambda_c I + J(x_c)^T J(x_c)$ is positive definite. The parameter λ is the Levenberg-Marquardt parameter (see (3.24)) and in order to obtain a more stable method, we combine the Levenberg-Marquardt method with Armijo rule [27, 59].

We therefore obtain the following theorem:

Theorem 3.3.2. *Let J be Lipschitz continuous and x_k be the Levenberg-Marquardt-Armijo iterates. Suppose that $||J(x_k)||$ is uniformly bounded and that the sequence of the Levenberg-Marquardt parameters λ_k is such that*

$$\lambda_k I + J(x_k)^T J(x_k) \tag{3.37}$$

is bounded. Then

$$\lim_{k\to\infty} J(x_k)^T R(x_k) = 0 \tag{3.38}$$

and for any limit point x^* of $\{x_k\}$ at which $R(x^*) = 0$, $J(x_k)$ has full column rank and $\lambda_k \to 0$, then $x_k \to x^*$ q-superlinearly. If

$$\lambda_k = O(||R(x_k)||) \tag{3.39}$$

as $k \to \infty$ then the convergence is q-quadratic.

Proof. The interested reader referred to the book [40]. □

3.4 Qualitative Analysis of the Model Problem

This section deals with the theoretical investigation of the ordinary differential equations $\frac{dN_S}{dt} = f(N_S)$ used for the description of the development of the occupied surface states, stated in Section 2.3.4.

We will prove the Theorem of Picard-Lindelöf, which gives the uniqueness and existence of a solution of the differential equation $\dot{\mathbf{x}} = f(t, \mathbf{x})$, and state then the assumptions on the coefficients of our model equations in order to apply the theorem.

3.4.1 Uniqueness and Existence of the solution

In the following, the Theorem of Picard-Lindelöf is stated, which essentially uses the Banach fixed-point Theorem.

Theorem 3.4.1. ***Banach fixed-point Theorem*** *Let $f : D \subset \mathbb{R}^d \to D \subset \mathbb{R}^d$, D complete with respect to the Euclidean metric, be a contraction, i.e., there exists a constant $c \in [0, 1)$ such that*

$$d(f(x), f(y)) \leq c d(x, y) \tag{3.40}$$

for all $x, y \in D$. Then has f a unique fixed point.

Proof. The interested reader is referred to one of the books [19, 66, 11]. □

Theorem 3.4.2. ***Theorem of Picard-Lindelöf*** *Let the function $\boldsymbol{f}(x, \boldsymbol{y})$ on the domain $D := J \times \mathbb{R}^n$, $J = [x_0, x_0 + a]$ be continuous and fulfil in D a Lipschitz-condition with respect to $\boldsymbol{y}$, i.e., there exists a constant L such that*

$$|\boldsymbol{f}(x, \boldsymbol{y}) - \boldsymbol{f}(x, \bar{\boldsymbol{y}})| \leq L|\boldsymbol{y} - \bar{\boldsymbol{y}}| \quad \text{for all } (x, \boldsymbol{y}), (x, \bar{\boldsymbol{y}}) \in D. \tag{3.41}$$

Then the initial value problem (IVP)

$$\boldsymbol{y}' = \boldsymbol{f}(x, \boldsymbol{y}), \quad \boldsymbol{y}(x_0) = \boldsymbol{y}_0 \tag{3.42}$$

has a unique solution for all $(x_0, \boldsymbol{y}_0) \in D$[78].

Proof. The idea behind the proof is to reformulate the IVP into the form $\mathbf{y} = T\mathbf{y}$. Then to show that T is a contraction on an appropriate interval. Finally, use the Banach fixed-point Theorem for uniqueness as well as for existence of the IVP because the solutions of the equation $\mathbf{y} = T\mathbf{y}$ are exactly the solutions of the IVP.

We start out by using the fundamental theorem of Calculus and so

$$\mathbf{y}' = \mathbf{f}(x, \mathbf{y}), \quad \mathbf{y}(x_0) = \mathbf{y}_0 \tag{3.43}$$

implies that

$$\mathbf{y}(x) = \mathbf{y}_0 + \int_{x_0}^{x} \mathbf{f}(t, \mathbf{y}(t))dt, \tag{3.44}$$

where $x \in J = [x_0, x_0+a]$. The equation $\mathbf{y}' = \mathbf{f}(x, \mathbf{y})$ implies that $\mathbf{y}(x)$ is continuously differentiable for $x \in J$. On the other hand, continuous solutions of (3.44) in J solve the IVP $\mathbf{y}(x_0) = \mathbf{y}_0$. Therefore, (3.42) and (3.44) are equivalent.

Therefore, we can write the IVP also in the form

$$\mathbf{y} = T\mathbf{y} \text{ with } (T\mathbf{y})(x) := \mathbf{y}_0 + \int_{x_0}^{x} \mathbf{f}(t, \mathbf{y}(t))dt \tag{3.45}$$

and so the integral operator T assigns to each function $\mathbf{y} \in C(J)$ a function $T\mathbf{y}$ from $C(J)$.

Next, we want to apply the Banach fixed-point Theorem in order to show that T has a unique fixed point that equals the unique solution of the IVP. Therefore, we need to show that T suffices a Lipschitz condition with Lipschitz constant $q < 1$.

Norming the space $C(J)$ with the maximum norm $||\mathbf{y}|| = \max_J e^{-2Lx}|\mathbf{y}(x)|$, where $|\mathbf{y}(x)| = \max_{i=1,\dots,n} |y_i(x)|$, we obtain from (3.41)

$$\begin{aligned} ||(T\mathbf{y})(x) - (T\mathbf{z})(x)|| = |\int_{x_0}^{x} [\mathbf{f}(t, \mathbf{y}(t)) - \mathbf{f}(t, \mathbf{z}(t))]dt| &\leq \int_{x_0}^{x} L\left(|\mathbf{y}(t) - \mathbf{z}(t)|e^{-2Lt}\right) e^{2Lt}dt \\ &\leq L\,||\mathbf{y} - \mathbf{z}|| \int_{x_0}^{x} e^{2Lt}dt \leq L\,||\mathbf{y} - \mathbf{z}||\,\frac{e^{2Lx}}{2L} \text{ for } y, z \in C(J). \end{aligned} \tag{3.46}$$

From (3.46) we have

$$\max_{x\in J} |(T\mathbf{y})(x) - (T\mathbf{z})(x)|e^{-2Lx} = ||(T\mathbf{y}) - (T\mathbf{z})|| \leq L\,||\mathbf{y} - \mathbf{z}||\,\frac{1}{2L} = \frac{1}{2}\,||\mathbf{y} - \mathbf{z}||\,. \tag{3.47}$$

Summarizing we have now shown that the operator T is a contraction with contraction constant $q = \frac{1}{2}$. From the Banach fixed-point Theorem follows now that the operator has a unique fixed point and so the IVP (3.42) has a unique solution.

□

Using the above theorem, we can now prove the existence and uniqueness of a solution of the ODE (4.3).

Corollary 3.4.3. *Suppose $f : D \subset \mathbb{R} \to \mathbb{R}$ given by (4.3), where the parameters ω_i are all non-negative, the occupied surface state density function N_S is C^1 and bounded for all $t \in [0, K]$ with $K > 0$. Then the ODE*

$$\frac{dN_S}{dt} = f(N_S) \tag{3.48}$$

has a unique solution for the initial value $N_S(0) = \rho_0$.

Proof. First, we calculate

$$f'(N_S) = -\frac{2AN_S}{T}\exp\left(-\frac{N_S^2+\omega_1}{T}\right) + \frac{2B}{T}\exp\left(-\frac{N_S^2+\omega_1}{T}\right)N_S^2 \\ -B\exp\left(-\frac{N_S^2+\omega_1}{T}\right) - C\exp\left(-\frac{\omega_2}{T}\right) - D\exp\left(-\frac{\omega_4}{T}\right) \tag{3.49}$$

and set $\overline{N_S} = ||N_S||_\infty$.We find

$$|f'(N_S)| \leq C_1\left|\frac{2N_S}{T}\right| + C_2\left|\frac{2N_S^2}{T}\right| + C_3 \leq L := C_1\left|\frac{2\overline{N_S}}{T}\right| + C_2\left|\frac{2\overline{N_S}^2}{T}\right| + C_3 \tag{3.50}$$

where

$$C_1 := \left|\frac{2A}{T}\right| \geq \left|\frac{2A}{T}\exp\left(-\frac{N_S^2+\omega_1}{T}\right)\right|$$
$$C_2 := \left|\frac{2B}{T}\right| \geq \left|\frac{2B}{T}\exp\left(-\frac{N_S^2+\omega_1}{T}\right)\right|$$
$$C_3 := |B+C+D| \geq \left|B\exp\left(-\frac{N_S^2+\omega_1}{T}\right) + C\exp\left(-\frac{\omega_2}{T}\right) + D\exp\left(-\frac{\omega_4}{T}\right)\right|$$

From the mean value theorem we obtain that

$$|f(N_S^1) - f(N_S^2)| \leq L|N_S^1 - N_S^2|.$$

Hence f satisfies a Lipschitz condition on D and therefore applying Theorem 3.4.2 shows that the solution exists and is unique. □

3.5 Poisson-Boltzmann Equation

The Poisson-Boltzmann equation is a partial differential equation that describes the electrostatic interactions between the molecules in ionic solutions. Ionic solutions are substances that contain free ions. In particular, ionic solutions are a subgroup of electrolytes and comprise acids, bases, and gases. The most common example for an ionic solution is salt placed in water, see Figure 3.3. There the individual components then dissociate due to the thermodynamic interactions between solvent and solute molecules [64].

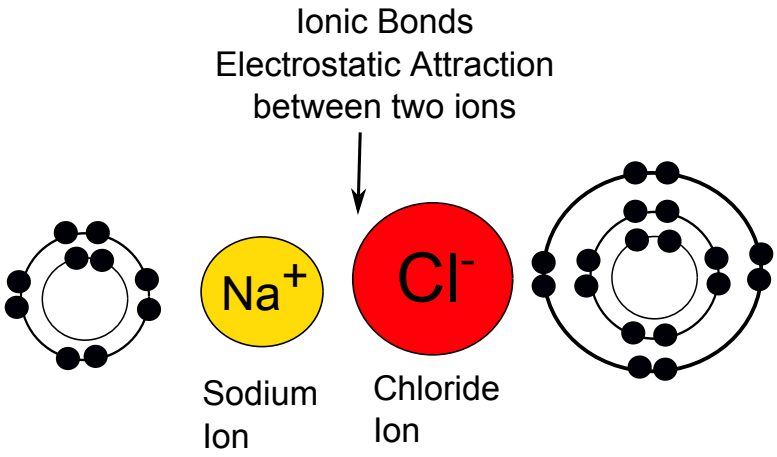

Figure 3.3: Dissolution of water.

3.5.1 Theory of the Poisson-Boltzmann Equation

The Poisson-Boltzmann equation represents the base for the Gouy-Chapman double layer (interfacial) theory and is also used in the modelling of implicit solvation, which approximates the effects of the solvent on the structures and interactions of proteins, DNA, RNA and other molecules in solutions of different ionic strengths. [16, 24]

The *ionic strength* I of a solution is a function of the concentration of all ions present in a solution by

$$I = \frac{1}{2}\sum_{j=1}^{p} c_i q_i^2, \tag{3.51}$$

where c_i is the molar concentration of ion i (dimension: mol m^{-3}) and q_i is the charge number of ion i.

3.5.2 Derivation of the Poisson-Boltzmann Equation from Debye-Hückel Theory

We split a domain Ω into three domains $\Omega_{int}, \Omega_{cen}$ and Ω_{out}. Ω_{int} contains the molecule for which we want to obtain the electrostatic potential. The outer region Ω_{out} consists of the solvent with the dielectric constant ϵ_{out} where we assume that the region contains mobile ions. The centered region Ω_{cen} is an exclusion layer around the molecule in Ω_{int} where no mobile charges of the solvent are present, but it has the same dielectric constant as region Ω_{out}, i.e. $\epsilon_{cen} = \epsilon_{out}$. Further, some solvent will also penetrate the region Ω_{int}, so we have the non-unit dielectric constant ϵ_{int}. For details about the specific assumptions and interrelations, the interested reader may refer to [53, 18, 74].

Furthermore we assume that the electrostatic potential satisfies Gauss's Law in all three regions, i.e.

$$\nabla \cdot \mathcal{E} = \frac{\rho}{\epsilon} \tag{3.52}$$

where $\nabla \cdot \mathcal{E}$ is the divergence of the electric field $\mathcal{E}$, ϵ is the permittivity and ρ is the charge density [35]. Assuming that the electric field is conservative, i.e. $\nabla \times \mathcal{E} = 0$, we are able to rewrite the equation in the form

$$\Delta V_{reg}(\mathbf{x}) = \frac{-4\pi\rho_{reg}(\mathbf{x})}{\epsilon_{reg}}$$

for each region Ω_{reg} where reg = inn, cen, out [31, 18].

If the molecule is represented by a series of N charges q_i at positions $\mathbf{x} \in \mathbb{R}^d$, $d = 1, 2, 3$ where $q_i = y_i q$, $y_i \in \mathbb{R}, i = 1, \ldots, N$, then the potential in region Ω_{int} is given by

$$V_{inn}(\mathbf{x}) = \sum_{i=1}^{N} \frac{q_i}{\epsilon_{int}|\mathbf{x} - \mathbf{x}_i|}.$$

By applying the Laplacian on both sides of the equation, we then have

$$\Delta V_{inn}(\mathbf{x}) = \sum_{i=1}^{N} \frac{-\pi q_i}{\epsilon_{int}} \delta(\mathbf{x} - \mathbf{x}_i),$$

where $\delta(\mathbf{x})$ is the *Dirac delta distribution*. For the centric region Ω_{cen} we obtain

$$\Delta V_{cen}(\mathbf{x}) = 0, \tag{3.53}$$

where we assume that there are no mobile ions implying the density charge is zero.

An assumption of the Debye-Hückel theory is that the ratio of the concentration of one type of ion near the molecule in Ω_{int} to its concentration far from Ω_{int} is given by the Boltzmann distribution law

$$e^{-\frac{W_+(\mathbf{x})}{k_B T}} \text{ resp. } e^{-\frac{W_-(\mathbf{x})}{k_B T}}, \tag{3.54}$$

where k_B is the Boltzmann constant, T is the temperature and $W_+(\mathbf{x}) = qV_{out}(\mathbf{x})$ resp. $W_-(\mathbf{x}) = -qV_{inn}(\mathbf{x})$ is the work required to move an ion of type + or − from $|\mathbf{x}| = \infty$ ($\lim_{|\mathbf{x}|\to\infty} V(\mathbf{x}) = 0$) to the point $\mathbf{x}$.

The Boltzmann distribution then yields

$$C_+ = Ce^{-\frac{qV_{out}(\mathbf{x})}{k_B T}} \text{ resp. } C_- = Ce^{+\frac{qV_{out}(\mathbf{x})}{k_B T}}$$

where we assume that $C = C_+ = C_-$ far from Ω_{int}.

We also obtain in the case of an $1:1$ electrolyte the charge density at a point x in Ω_{out} by

$$\rho_{out}(\mathbf{x}) = C_+ q - C_- q = Ce^{-\frac{qV_{out}(\mathbf{x})}{k_B T}} - Ce^{+\frac{qV_{out}(\mathbf{x})}{k_B T}} = -2Cq \sinh\left(\frac{qV_{out}(\mathbf{x})}{k_B T}\right).$$

In total, we get

$$\Delta V_{out}(\mathbf{x}) = -\frac{4\pi\rho_{out}(\mathbf{x})}{\epsilon_{out}} = \left(\frac{8\pi Cq}{\epsilon_{out}}\right) \sinh\left(\frac{qV_{out}(\mathbf{x})}{k_B T}\right) \tag{3.55}$$

and relating C to the Debye-Hückel parameter c, see [14], we can then rewrite equation (3.55) in the form

$$\Delta V_{out}(\mathbf{x}) = c^2 \left(\frac{k_B T}{q}\right) \sinh\left(\frac{q V_{out}(\mathbf{x})}{k_B T}\right). \tag{3.56}$$

From a physical perspective, we expect the function $V(\mathbf{x})$ at the interfaces of the regions as well as the dielectric constant times the normal derivative , $\epsilon \nabla V(\mathbf{x}) \cdot n$, of the function to be continuous where n is the unit outer normal vector [53].

From these assumptions we then obtain the following interface conditions: On $\partial\Omega_{inn:cen} := \Omega_{int} \cap \Omega_{cen}$, we have

$$V_{inn}(\mathbf{x}) = V_{cen}(\mathbf{x}), \quad \epsilon_{int} \nabla V_{inn}(\mathbf{x}) \cdot n = \epsilon_{cen} \nabla V_{cen}(\mathbf{x}) \cdot n \tag{3.57}$$

and on $\partial\Omega_{cen:out} := \Omega_{cen} \cap \Omega_{out}$ we obtain

$$V_{cen}(\mathbf{x}) = V_{out}(\mathbf{x}), \quad \epsilon_{cen} \nabla V_{cen}(\mathbf{x}) \cdot n = \epsilon_{out} \nabla V_{out}(\mathbf{x}) \cdot n. \tag{3.58}$$

As above we assume that $V(\infty) = 0$.

From these considerations we obtain the *nonlinear Poisson-Boltzmann equation*:

Definition 3.5.1. The **nonlinear Poisson-Boltzmann equation** (PBE) is a second order semi-linear elliptic equation

$$-\nabla \cdot (A \nabla V) + b(\mathbf{x}, V) = f \text{ in } \Omega, \tag{3.59a}$$

$$V = h_D \text{ on } \Gamma_D, \tag{3.59b}$$

$$(A \nabla V) \cdot \mathbf{n} + cV = h_N \text{ on } \Gamma_N, \tag{3.59c}$$

where $\Omega \subset \mathbb{R}^d$ is a bounded region with boundary $\Gamma = \Gamma_D \cup \Gamma_N$ and $\Gamma_D \cap \Gamma_N = \emptyset$, $b(\mathbf{x}, V(\mathbf{x})) : \Omega \times \mathbb{R} \to \mathbb{R}$ and $b = c(\mathbf{x})^2 \sinh(V(\mathbf{x})) \in L^2(\Omega)$, $f = C \sum_{i=1}^{N} z_i \delta(\mathbf{x} - \mathbf{x}_i) \in L^2(\Omega)$ where $z_i \in [-1, 1]$ and $A : \mathbb{R}^d \to \mathbb{R}^d \times \mathbb{R}^d, A = \epsilon(\mathbf{x}) I$ is uniformly elliptic and bounded, with coercivity factor γ and bound constant C.

We will prove a uniqueness result for the solution of the nonlinear Poisson-Boltzmann equation and, first need some definitions and theorems.

Definition 3.5.2. Gateaux-derivative The mapping $F : \mathcal{D} \subset \mathcal{H}_1 \to \mathcal{H}_2$ is called Gateaux- or G-differentiable at $u \in \text{int}(D)$ if there exists $F'(V) \in \mathcal{L}(\mathcal{H}_1, \mathcal{H}_2)$, called the **G-derivative**, such that

$$\lim_{s \to 0} \frac{||F(V + sh) - F(V) - sF'(V)(h)||}{s} = 0 \tag{3.60}$$

holds for any $h \in \mathcal{H}_1$. In particular, the linear operator $F'(V)$ is unique [20].

Definition 3.5.3. The mapping $F : \mathcal{D} \subset \mathcal{H}_1 \to \mathcal{H}_2$ is called Frechet- or F-differentiable at $V \in \text{int}(D)$ if there exists $F'(V) \in \mathcal{L}(\mathcal{H}_1, \mathcal{H}_2)$, called the **F-derivative**, such that

$$\lim_{||h|| \to 0} \frac{||F(V + h) - F(V) - F'(V)(h)||}{||h||} = 0 \tag{3.61}$$

holds for any $h \in \mathcal{H}_1$. Again, the linear operator $F'(V)$ is unique. [20]

From the above definition it is clear that every F-differentiable function is also G-differentiable. However, the reverse is not true: consider a Banach space B and a linear functional f defined on B, which is discontinuous at $x = 0$. Set $F(x) = ||x||\, f(x)$. Then $F(x)$ is Gateaux-differentiable at $x = 0$ with derivative 0, because

$$\frac{||F(0+sh) - F(0) - sF'(0)(h)||}{s} = \frac{||\, ||sh||\, f(sh)||}{s} = |s|\, ||h||\, ||f(h)|| \to 0 \quad \text{for } s \to 0,$$

but $F(x)$ is not Frechet-differentiable as the limit

$$\lim_{||x|| \to 0} f(x)$$

does not exist [56].

The aim now is to prove the uniqueness and existence of the nonlinear Poisson-Boltzmann equation (3.59a) via the minimization of a functional $J : \mathcal{H} \to \mathbb{R}$. In order to refer to the Ekland-Temam Theorem, that returns the existence of a local minimizer, we need to impose some conditions on the functional, defined below:

The first condition is that of a proper functional, which means that the functional is not too degenerated.

Definition 3.5.4. proper functional A functional $J(\cdot)$ is called *proper* if $J \not\equiv +\infty$ and $J(u) > -\infty$ for all $u \in \mathcal{D}$.

The second condition is coercivity, which means that the functional cannot take on a minimizer when it is very far away from the origin.

Definition 3.5.5. coercive functional A functional $J(\cdot)$ is called *coercive* if $J(u) \to \infty$ when $||u|| \to +\infty$.

The next definition revolves around lower semi-continuity, which turns out to be essential in order for the minimum to be taken on by the to be minimized function.

Definition 3.5.6. lower semi-continuity A functional $J(\cdot)$ is called *lower semi-continuous* at the point $u \in \mathcal{D}$ if for all $s < J(u)$ there exists $\epsilon > 0$ such that the inequality $s < J(v)$ holds for all v with $||v - u|| < \delta$.

An equivalent condition is, see [37],

$$J(u) = \liminf_{v \to u} J(v) = \sup_{\epsilon > 0} \inf\{J(v) |\, ||v - u|| < \epsilon\}. \tag{3.62}$$

Next we introduce the notion of optimality, where we need to distinguish between the global property and a local property [22].

Definition 3.5.7. Global minimizer A point $V^* \in \mathcal{H}$ such that $J(V^*) = \min_{V \in \mathcal{H}} J(V)$ is called a *global minimizer*.

Definition 3.5.8. Local minimizer A point $V^* \in \Omega \subset \mathcal{H}$ is called a *local minimizer* if $J(V^*) = \min_{V \in \Omega} J(V)$.

Important for relating the PBE to the functional $J(\cdot)$, we define the gradient mapping, which states the relation of a minimizer V^* and the value of the G-derivative of $J(\cdot)$ at V^*.

Definition 3.5.9. gradient mapping The mapping $F : \mathcal{D} \subset \mathcal{H} \to \mathcal{H}$ is called *gradient mapping* if for a G-differentiable functional $J : \mathcal{D} \subset \mathcal{H} \to \mathbb{R}$ the equation $F(V) = J'(V)$ holds for all $V \in \mathcal{D}$.

In particular to ensure that a local minimizer is also a global minimizer, we need the concept of convexity.

Definition 3.5.10. Convexity The functional $J : \mathcal{D} \subset \mathcal{H} \to \mathbb{R}$ is called *convex* if the inequality

$$J(\lambda U + (1-\lambda)V) \leq \lambda J(U) + (1-\lambda)J(V) \tag{3.63}$$

holds for all $V, U \in \mathcal{D}$ and $\lambda \in (0,1)$ provided the right-hand side is defined.

Building on these definitions, we are able to state the Euler condition that formalizes the relationship between minimizers and the gradient mapping.

Theorem 3.5.11. ***Euler condition*** *For a G-differentiable functional $J : \mathcal{D} \subset \mathcal{H} \to \mathbb{R}$ with $F(V) = J'(V)$ for all $V \in \mathcal{D}$ and a local minimizer V^* of $J(\cdot)$, the equation $F(V^*) = 0$ holds.*

Proof. The interested reader may refer to the book [19]. □

Theorem 3.5.12. ***Ekland-Temam Theorem*** *If $J : \mathcal{D} \subset \mathcal{H} \to \mathbb{R}$ is a convex, lower semicontinuous, proper and coercive functional and $\mathcal{D}$ a non-empty closed convex subset of $\mathcal{H}$, then $J(\cdot)$ has a local minimizer $V^* \in \mathcal{D}$.*

Proof. For a proof see Proposition 1.2 in [22]. □

Theorem 3.5.13. *The nonlinear Poisson-Boltzmann equation (3.59a) has a unique weak solution $V \in H^1(\Omega)$.*

Proof. The idea is to transform the nonlinear PB equation (3.59a) into its weak form

$$\mathcal{A}(V, v) + (N(V), v) = F(v) \quad \forall v \in H_0^1(\Omega) \tag{3.64}$$

where

$$\mathcal{A}(V, v) = \int_\Omega A\nabla V \cdot \nabla v d\mathbf{x}, \tag{3.65a}$$

$$(N(V), v) = \int_\Omega b(\mathbf{x}, w + V) v d\mathbf{x}, \tag{3.65b}$$

$$F(v) = \int_\Omega f v d\mathbf{x} - A(w, v) \tag{3.65c}$$

and the fixed known function $w \in H^1(\Omega)$ has trace $h = \operatorname{tr} w$. Then we use the representation (3.64) to define a convex functional $J(.)$ and minimize it over the domain. The minimizer of the functional $J(.)$ satisfies, due to the Euler condition (Theorem 3.5.11), $J'(V^*) = 0$ and $J'(V^*) = \mathcal{A}(V, v) + (N(V), v) - F(v)$. But this implies that the minimum is a solution of the nonlinear

PB equation. The uniqueness follows from showing the strict convexity of $J(.)$, implying that the minimizer V^* is unique and therefore the solution as well.

We start out by defining

$$r(\mathbf{x}, y) := \frac{c(\mathbf{x})^2}{2}\left[(e^y - 1)e^{w(\mathbf{x})} + (e^{-y} - 1)e^{-w(\mathbf{x})}\right]. \tag{3.66}$$

In particular $r(\mathbf{x}, 0) = 0$ and

$$r'(\mathbf{x}, y) := \frac{\partial r(x, y)}{\partial y} = \frac{c^2(\mathbf{x})}{2}(e^y e^{w(\mathbf{x})} - e^{-y}e^{-w}) = c^2(\mathbf{x})\sinh(w(\mathbf{x}) + y) = b(\mathbf{x}, w + y).$$

Defining the functional $J(\cdot) : H_0^1(\Omega) \to \mathbb{R}$ as

$$J(V) := \begin{cases} \frac{1}{2}\mathcal{A}(V, V) + \int_\Omega r(\mathbf{x}, V)d\mathbf{x} - F(V), & r(\cdot, V) \in L^2(\Omega) \text{ for } V \in H_0^1(\Omega), \\ +\infty & r(\cdot, v) \notin L^2(\Omega) \text{ for } V \in H_0^1(\Omega) \end{cases} \tag{3.67}$$

we will show that the Frechet-derivative of $J(\cdot)$ satisfies for a local minimizer $J'(V^*) = 0$ equation (3.67) and V^* solves the equation

$$(J'(V), v) = 0, \quad \text{for all } v \in H_0^1(\Omega). \tag{3.68}$$

We can rewrite $J(\cdot)$ in the form

$$J(V) = \frac{1}{2}\mathcal{A}(V, V) + (r(\mathbf{x}, V), 1) - F(V)$$

and consider

$$\begin{aligned} J(V + h) - J(V) = \left(\frac{1}{2}\mathcal{A}(V + h, V + h) + (r(\mathbf{x}, V + h), 1) - F(V + h)\right) \\ -[\frac{1}{2}\mathcal{A}(V, V) + (r(\mathbf{x}, V), 1) - F(V))] = \frac{1}{2}[\mathcal{A}(V, h) + \mathcal{A}(h, V) + \mathcal{A}(h, h)] + \\ \big((r(\mathbf{x}, u) + r'(\mathbf{x}, u)h + \ldots, 1) - (r(\mathbf{x}, V), 1)\big) + (F(V) + F(h) - F(V)) \\ = \mathcal{A}(V, h) + (r'(\mathbf{x}, V), h) - F(h) + 0(||h||^2). \end{aligned}$$

From this we obtain the Euler condition

$$\text{find } V \in \mathcal{H} \text{ such that } (J'(V), v) = \mathcal{A}(V, v) + (N(V), v) - F(v) = 0 \quad \forall v \in \mathcal{H}, \tag{3.69}$$

where $N(V) = \frac{\partial R(\mathbf{x}, V)}{\partial y} = b(\mathbf{x}, w+V)$ represents the weak formulation of the nonlinear PB equation.

For showing that $J(\cdot)$ has a minimum $V \in H_0^1(\Omega)$, we will apply the Ekland-Temam-Theorem (Theorem 3.5.12) which returns a local minimizer of the functional J. To apply the theorem, we need to show that $J(\cdot)$ is proper, convex, lower semi-continuous and coercive on $H_0^1(\Omega)$.

From the definition of $J(\cdot)$ and the corresponding definition of $R(\mathbf{x}, y)$, we obtain that

$$J(V) > -\infty, \quad \text{for all } V \in H_0^1(\Omega).$$

J is not identically equal to $+\infty$ because, e.g., for $V = 0$ we have $\mathcal{A}(0,0) = 0, r(\mathbf{x}, 0) = 0$, and $F(0) = 0$ yielding $J(0) = 0$. Therefore J is a proper functional.

Next, we will show that $J(\cdot)$ is convex. First F is a linear functional, and from linearity

$$F(\lambda V + (1-\lambda)U) = \lambda F(V) + (1-\lambda)F(U)$$

it follows that $-F$ is trivially convex. From the uniform ellipticity of A we can deduce that $\mathcal{A}^*(V) = \mathcal{A}(V, V)$ is convex. This follows from

$$J(\theta) = \theta^T A(x)\theta \geq \alpha|\theta|^2$$

and so

$$J''(\theta) = A(x) \geq \alpha > 0,$$

which implies the convexity of J and in particular of $\mathcal{A}^*$.

Therefore in order for $J(\cdot)$ to be convex, we need to show that $r(\mathbf{x}, y)$ is convex and the convexity of $r(\mathbf{x}, y)$ is equivalent to the condition that if $r :\to \mathbb{R}$ is G-differentiable on a convex set C,

$$(r'(\mathbf{x}, V) - r'(\mathbf{x}, U))(V - U) \geq 0 \quad \text{for all } V, U \in K, \text{ for all } \mathbf{x} \in \Omega$$

From

$$r''(\mathbf{x}, V) = c^2(\mathbf{x})\cosh(w(\mathbf{x}) + V) \geq 0, \quad \forall\, V \in K$$

we have that $r'(\mathbf{x}, V) = c^2(\mathbf{x})\sinh(w(\mathbf{x}) + V)$ is monotonically increasing. This implies that

$$\text{sign}(U - V) = \text{sign}(r'(\mathbf{x}, U) - r'(\mathbf{x}, V))$$

for all $x \in \Omega$ and $U, V \in K$ and so $r(\cdot, V)$ is convex for $V \in K$.

Now for $R(V) = \int_\Omega r(\cdot, V)d\mathbf{x}$ we obtain

$$\begin{aligned} R(\lambda V + (1-\lambda)U) &= \int_\Omega r(\mathbf{x}, \lambda V + (1-\lambda)U)\mathbf{x} \leq \int_\Omega [\lambda r(\mathbf{x}, V) + (1-\lambda)r(\mathbf{x}, U)]\mathbf{x} \\ &= \lambda \int_\Omega r(\mathbf{x}, V)d\mathbf{x} + (1-\lambda)\int_\Omega r(\mathbf{x}, U)d\mathbf{x} = \lambda R(V) + (1-\lambda)R(U) \end{aligned}$$

which implies that $R(V)$ is also convex as a function of V. So $J(\cdot)$ is a linear combination of convex functions and therefore is itself convex on K. The proof of the lower semi-continuity can be obtained from [33]. For the coerciveness of $J(\cdot)$, we see that we obtain for $\alpha = \inf_\Omega w$ and $\beta = \sup_\Omega w$

$$r(\mathbf{x}, V) = \frac{c^2}{2}[(e^V - 1)e^w + (e^{-V} - 1)e^{-w}] \geq -\frac{c^2}{2}[e^\beta + e^{-\alpha}] > -\infty$$

and so

$$\int_\Omega r(\mathbf{x}, V)d\mathbf{x} \geq -\text{meas}(\Omega)\frac{c^2}{2}[e^\beta + e^\alpha] > -\infty.$$

Then we obtain from the uniform ellipticity and boundedness of the function $\epsilon(\mathbf{x})$ that

$$\sqrt{\gamma}\,||V||_{H^1(\Omega)} \leq ||V||_A \leq \sqrt{C}\,||V||_{H^1(\Omega)}$$

for $||V||_A = \mathcal{A}(V,V)^{1/2} = \left(\int_\Omega A\nabla V \cdot \nabla V d\mathbf{x}\right)^{1/2}$.

Next we estimate

$$J(V) = \frac{1}{2}\int_\Omega A\nabla V \cdot \nabla V d\mathbf{x} + \int_\Omega r(\mathbf{x},V)d\mathbf{x} - F(V)$$

$$\geq \frac{\gamma}{2}\,||V||^2_{H^1(\Omega)} - \text{meas}(\Omega)\frac{c^2}{2}[e^\beta + e^{-\alpha}] - ||f||_{L^2(\Omega)}\,||V||_{L^2(\Omega)} - C\,||w||_{H^1(\Omega)}\,||V||_{H^1(\Omega)}$$

and as f respectively w are fixed and

$$\lim_{||V||_{L^2(\Omega)}\to\infty} \frac{||V||^2_{H^1(\Omega)}}{||V||_{L^2(\Omega)}} = \lim_{||V||_{L^2(\Omega)}\to\infty} \frac{||V||^2_{L^2(\Omega)} + |V|^2_{H^1(\Omega)}}{||V||_{L^2(\Omega)}} \geq \lim_{||V||_{L^2(\Omega)}\to\infty} ||V||_{L^2(\Omega)} = +\infty$$

holds, we obtain $J(V) \to +\infty$ for $||V||_{H^1(\Omega)} \to +\infty$, showing that $J(\cdot)$ is coercive on $H^1_0(\Omega)$.

The Ekland-Temam theorem (Theorem 3.5.12) now yields that there exists a minimizer $V \in H^1_0(\Omega)$ of $J(\cdot)$.

We have still to prove the uniqueness of the minimizer of $J(\cdot)$ and therefore the uniqueness of the solution of the nonlinear PBE.

Assume that V_1 and V_2 are two solutions of (3.64) and define $V^- = V_1 - V_2 \in H^1_0(\Omega)$. We obtain

$$\mathcal{A}(V_1 - V_2, v) + (N(V_1) - N(V_2), v) = 0 \quad \forall v \in H^1_0(\Omega)$$

and so from the monotonic increase of $N(V)$ in V, we find that $(N(V_1) - N(V_2), V_1 - V_2) \geq 0$ for $v = V_1 - V_2$ and so in order to satisfy (3.70) we must have $\mathcal{A}(V_1 - V_2, V_1 - V_2) \leq 0$.

On the other side, as A is uniformly elliptic, we have

$$\mathcal{A}(V_1 - V_2, V_1 - V_2) = ||V_1 - V_2||^2_A \geq 0.$$

Therefore, both equations can only hold if

$$||V_1 - V_2||_A = 0$$

and that implies $V_1 = V_2$.

□

3.5.3 Poisson-Boltzmann Equation for Gas Sensor Simulations

In the following, we will discuss some of the properties originating from the version of the nonlinear Poisson-Boltzmann equation considered for the nanowire gas sensor simulations. In the previous section we have provided a proof for the general PBE via minimizing a functional. For this approach we needed some theory from calculus of variations and optimization, in particular the Ekland-Temam Theorem (Theorem 3.5.12) and the Euler condition (Theorem 3.5.11).

In the following section, we will provide a different proof for the PBE equation for gas sensors, that uses the Leray-Schauder fixed-point Theorem together with an estimate for semilinear elliptic equations. The advantage of this approach is that it does not require the deep theory of the calculus of variation and provides a deeper insight in the specific form of the PBE for nanowire gas sensors.

The **nonlinear Poisson-Boltzmann equation** for the use for gas sensors is the BVP

$$-\nabla \cdot (\epsilon \nabla V_e) + q n_i (e^{V_e/U_T} - e^{-V_e/U_T}) - qC = 0 \text{ in } \Omega \tag{3.70a}$$

$$\nabla_n V_e = 0 \text{ on } \partial\Omega_N \tag{3.70b}$$

$$V_e = V_D \text{ on } \partial\Omega_D \tag{3.70c}$$

where V_e is the equilibrium potential, U_T is the thermal Voltage, q is the elementary charge, n_i intrinsic carrier density, $C = N_d^+ - N_a^- + n_S$ where N_d^+ is the donor concentration, N_a^- is the acceptor concentration and n_S the concentration of electrons from the surface equation.

Our goal is now to prove uniqueness for a solution of the above equation, therefore ensuring that in the numerical calculation the solution does not oscillate between two or more solutions. We need to start with an estimate for semilinear elliptic equations, then state the Leray-Schauder fixed-point Theorem and finally we are able to prove uniqueness for the solution of equation (3.70a).

Theorem 3.5.14. *Let $v(x, y)$ be a solution of*

$$F(x, y, u, v_x, v_y, v_{xx}, v_{xy}, v_{yy}) = f(x, y) \text{ in } D$$
$$v = h \text{ on } \partial D.$$

Let w and W satisfy the inequalities

$$F(x, y, W, W_x, W_y, W_{xx}, W_{xy}, W_{yy}) \leq f(x, y) \leq F(x, y, w_x, w_y, w_{xx}, w_{xy}, w_{yy}) \tag{3.71}$$

in D and

$$w(x, y) \leq h(x, y) \leq W(x, y) \quad \text{on } \partial D. \tag{3.72}$$

Assume that for each constant $\lambda \in [0, 1]$, the function F is elliptic with respect to $\lambda w + (1 - \lambda)v$ and $(1 - \lambda)v + \lambda W$ in D, and that $\partial F/\partial v \leq 0$ in D. Then we have

$$w(x, y) \leq v(x, y) \leq W(x, y) \quad \text{in } \partial D. \tag{3.73}$$

Proof. For a proof see the book [62], p.115. □

Next, important for the proof of the lemma below, we need to state the Leray-Schauder fixed-point Theorem, which plays a fundamental role in proving the existence of a solution in the lemma below.

Theorem 3.5.15 (Leray-Schauder fixed-point Theorem)**.** *Let B be a Banach space and T be a compact mapping from $B \times [0, 1]$ into B such that $T(x, 0) = 0$ for all $x \in B$. Furthermore, assume that there exists a constant $C > 0$ such that*

$$||x||_B < C \tag{3.74}$$

for all $(x, t) \in B \times [0, 1]$ with $x = T(x, t)$. Then the mapping T_1 from B into itself given by $T_1(x) = T(x, 1)$ has a fixed point [26].

Equipped with those two results, we are able to prove the following lemma.

Lemma 3.5.16. *Assume the following four conditions:*

1. *The function $h(x,\psi) \in C^1(\Omega \times \mathbb{R})$ is monotonically increasing in ψ for all $x \in \Omega$.*
2. *$a \in L^\infty(\Omega)$ and there exists a constant c such that $a(x) \geq c > 0$ holds for all $x \in \Omega$.*
3. *There exist functions $\underset{\sim}{h}(\psi)$ and $\tilde{h}(\psi)$ with*

$$\underset{\sim}{h}(\psi) \leq h(x,\psi) \leq \tilde{h}(\psi) \quad \forall x \in \Omega, \forall \psi. \tag{3.75}$$

4. *The algebraic equations $\underset{\sim}{h}(\tilde{\psi}) = 0$ and $\tilde{h}(\underset{\sim}{\psi}) = 0$ have solutions $\underset{\sim}{\psi}$ and $\tilde{\psi}$.*

Then there exists a unique solution ψ of the semilinear elliptic boundary-value problem

$$-\nabla \cdot (a(x)\nabla\psi) + h(x,\psi) = 0 \text{ in } \Omega, \tag{3.76a}$$
$$\nabla_n \psi = 0 \text{ on } \partial\Omega_N, \tag{3.76b}$$
$$\psi = \psi_D \text{ on } \partial\Omega_D, \tag{3.76c}$$

and $\psi \in H^1(\Omega) \cap L^\infty(\Omega)$. The solution ψ satisfies the estimate

$$\min(\inf_{\partial\Omega_D} \psi_D, \underset{\sim}{\psi}) =: \underline{\psi} \leq \psi(x) \leq \overline{\psi} := \max(\sup_{\partial\Omega_D} \psi_D, \tilde{\psi}), \quad \forall \psi \in \bar{\Omega}. \tag{3.77}$$

Proof. The proof of the lemma will proceed as follows:

1. Proof of uniqueness of solution.
2. Proof of estimate (3.77) if the solution exists.
3. Proof of existence.

(1) Suppose that $\psi_1, \psi_2 \in H^1(\Omega) \cap L^\infty(\Omega)$ are two solutions. Define $\phi = \psi_1 - \psi_2$, so ϕ solves the boundary value problem

$$-\nabla \cdot (a(x)\nabla\phi) + h(x,\psi_1) - h(x,\psi_2) = 0 \text{ in } \Omega, \tag{3.78a}$$
$$\nabla_n \phi = 0 \text{ on } \partial\Omega_N, \tag{3.78b}$$
$$\phi = 0 \text{ on } \partial\Omega_D. \tag{3.78c}$$

The equations can be rewritten as

$$-\nabla \cdot (a(x)\nabla\psi) + h(x,\psi_1) - h(x,\psi_2) = -\nabla \cdot (a(x)\nabla\psi) + \partial_\psi h(x,\hat{\psi}(x))\phi \tag{3.79}$$

by using the mean-value theorem

$$h(x,\psi_1) - h(x,\psi_2) = \partial_\psi h(x,\hat{\psi}(x))(\psi_1 - \psi_2).$$

Since h increases monotonically we can deduce that $\partial_\psi h(x, \hat{\psi}(x)) \geq 0$, so applying the weak maximum principle (see, e.g., [23]), which states that the solution takes on its maximum on the boundary, implies that $\phi = 0$. This shows the uniqueness of the solution.

(2) We apply now Theorem 3.5.14 to show that the estimate (3.77) is satisfied for a solution. Set

$$w := \underline{\psi}, \tag{3.80a}$$

$$W := \overline{\psi}. \tag{3.80b}$$

We now need to show that the function

$$F(x, y, z, z_x, z_y, z_{xx}, z_{xy}, z_{yy}) := -(-\nabla \cdot (a(x)\nabla\psi) + h(x, \psi))$$

is elliptic with respect to $\lambda\underline{\psi} + (1-\lambda)\psi$ and $\lambda\overline{\psi} + (1-\lambda)\psi$ where $\lambda \in [0, 1]$.

We obtain $z_1^2\partial_{s_1}F + z_1z_2\partial_{s_2}F + z_2^2\partial_{s_3}F = 2a(z_1, z_2)z_1^2 + 2a(z_1, z_2)z_2^2 \geq 2\underline{a}(z_1^2 + z_2^2) > 0$ for $z_1^2 + z_2^2 > 0$, since $\underline{a} > 0$ because a is uniformly elliptic.

This implies that $\partial F/\partial\psi \leq 0$ is satisfied due to the monotonicity of $h(x, \psi)$ with respect to ψ.

Next we need to show that $F(W) \leq 0 \leq F(w)$. For $\overline{\psi}$ and $\underline{\psi}$ constant we get

$$F(W) = F(\overline{\psi}) = -h(x, \overline{\psi}) \leq -h(x, \tilde{\psi}) \leq -\underline{h}(\tilde{\psi}) = 0, \tag{3.81a}$$

$$F(w) = F(\underline{\psi}) = -h(x, \underline{\psi}) \geq -h(x, \underset{\sim}{\psi}) \geq -\tilde{h}(\underset{\sim}{\psi}) = 0 \tag{3.81b}$$

and so $\overline{\psi} \geq \tilde{\psi}$ and $\underline{\psi} \geq \underset{\sim}{\psi}$.

From Theorem 3.5.14 we then obtain estimate (3.77).

(3) For $\psi \in L^2(\Omega)$ define the cut function

$$\psi_K := \begin{cases} -K, & \text{if } \psi(x) \leq -K, \\ \psi(x), & \text{if } -K \leq \psi(x) \leq K, \\ K, & \text{if } K \leq \psi(x), \end{cases} \tag{3.82}$$

where $K > 0$. For $\psi \in H^1(\Omega)$ ψ_K is in $H^1(\Omega)$ as well as in $L^\infty(\Omega)$. Define K as follows

$$K := \max(|\underline{\psi}|, |\overline{\psi}|).$$

Furthermore define the operator

$$M : L^2(\Omega) \times [0, 1] \to L^2(\Omega), \quad M(y, \sigma) = w, \tag{3.83}$$

where w is the solution of the linear elliptic boundary-value problem

$$-\nabla \cdot (a(x)\nabla w) + \sigma h(x, y_K) = 0 \text{ in } \Omega, \tag{3.84a}$$

$$\nabla_n w = 0 \text{ on } \partial\Omega_N, \tag{3.84b}$$

$$w = \sigma\psi_D \text{ on } \partial\Omega_D. \tag{3.84c}$$

Obviously, every fixed point ϕ_1 of $M(.,1)$, i.e. every ϕ_1 with

$$M(\phi_1, 1) = \phi_1 \tag{3.85}$$

that satisfies $|\phi_1| \leq K$ a.e. on Ω, is a weak solution of the original problem.

From the smoothness of f and $\partial\Omega$, we obtain (see [60], p. 49, and for the regularity of the solution of the elliptic BVP, [23])

$$\phi_1 \in C^1(\Omega).$$

Therefore the set

$$\Omega_+ := \{x \in \Omega | \phi_1 > K\} \subset \Omega$$

is open in Ω and the boundary $\partial\Omega_+$ consists of points x with either $\phi_1(x) = K$ or $x \in \partial\Omega$.

For Ω_+ non-empty and $x^* \in \Omega_+$, we define the maximal connected component of Ω_+ containing x^* by Ω_+^*. Then $\phi_1|_{\Omega_+^*}$ solves the boundary-value problem

$$\begin{aligned} -\nabla \cdot (a(x)\nabla\phi_1) + h(x, \phi_1) &= 0 \text{ in } \Omega_+^*, \\ \nabla_n \phi_1 &= 0 \text{ on } \partial\Omega_+^* \cap \partial\Omega_N, \\ \phi_1 &= \phi_D \text{ on } \partial\Omega_+^* \cap \partial\Omega_D, \\ \phi_1 &= K \text{ on } \partial\Omega_+^* \backslash \partial\Omega. \end{aligned}$$

From $\overline{\psi} \leq K$ and therefore $h(x, K) \geq 0$ due to $0 = \underset{\sim}{h}(\tilde{\psi}) \leq h(x, \tilde{\psi}) \leq h(x, K)$ and as $\sup_{\partial\Omega_D} \phi_D \leq K$, the constant $\bar{\phi_1} := K$ is an upper solution of the last problem. But this implies that $\phi_1 \leq K$ in $\bar{\Omega}$ and therefore Ω_+ is empty. In a similar way can be shown that $\phi_1(x) \geq -K$ in $\bar{\Omega}$.

Next, the operator

$$\begin{aligned} T : L^2(\Omega) &\to L^2(\Omega), \\ y &\mapsto y_K \end{aligned} \tag{3.86}$$

is continuous, because $||T(y)||_{L^2} \leq ||y||_{L^2}$ and so

$$||T(x) - T(y)||_{L^2} = ||x_K - y_K||_{L^2} \leq ||x - y||_{L^2}$$

holds due to the lower and upper bound.

From the fact that the right side of (3.84a) depends continuously in $L^2(\Omega)$ on $(y, \sigma) \in L^2(\Omega) \times [0, 1]$, that solutions of linear elliptic equations in $H^1(\Omega)$ depend continuously on right sides in $L^2(\Omega)$ and boundary data on $H^1(\Omega)$, we obtain that the operator M is continuous.

Next, it will be shown that the range of M is a bounded set in $H^1(\Omega)$. From regularity results (see again [60, 23]) for the elliptic problem (3.84a), we obtain

$$||w||_{H^1} \leq C(||\sigma h(., y_k)||_{L^2} + ||\sigma w_D||_{H^1}) \leq C(|\Omega|^{1/2} \sup_{x \in \bar{\Omega}} |h(x, K)| + ||w_D||_{H^1})$$

because h is bounded as a function of x. From the Rellich-Kondrachev compactness theorem, we know $H^1(\Omega) \subset\subset L^2(\Omega)$ and the operator M is compact because the range of M is bounded.

Finally, the Leray-Schauder fixed-point Theorem (Theorem 3.5.15, also [61]) yields the existence of a fixed point ϕ_1 of $M(., 1)$ and hence the existence of a solution of (3.76).

□

Building on the above lemma, we are now able to prove the unique solution of the Poisson-Boltzmann equation (3.70a) and an estimate for the solution.

Theorem 3.5.17. *Let $\Omega \subset \mathbb{R}^d$ be open, bounded, ϵ is uniformly elliptic and $C \in L^\infty(\Omega)$. Then the Poisson-Boltzmann equation (3.70a) has a unique solution $V_e \in H^1 \cap L^\infty(\Omega)$. Furthermore, the estimate*

$$\min(\inf_{\partial\Omega_D} V_D, U_T \sinh^{-1}(\frac{\inf_\Omega C}{2qn_i})) \leq V_e(x) \leq \max(\sup_{\partial\Omega_D} V_D, U_T \sinh^{-1}(\frac{\sup_\Omega C}{2qn_i})) \tag{3.87}$$

holds for all $x \in \Omega$.

Proof. Set $h(x,V) := qn_i(e^{V/U_T} - e^{-V/U_T}) - qC$. Furthermore

$$\partial_V h(x,V) = qn_i(e^{V/U_T} + e^{-V/U_T})/U_T = (2qn_i/U_T)\cosh(V/U_T) > 0 \tag{3.88}$$

holds for all $V \in \mathbb{R}$ and the estimate

$$\underset{\sim}{h}(V) := 2qn_i \sinh(V/U_T) - \sup_\Omega C \leq h(x,V) \leq 2qn_i \sinh(V/U_T) - \inf_\Omega C =: \tilde{h}(V) \tag{3.89}$$

Solving for $\underset{\sim}{h}(\tilde{V}) = 0$ and $\tilde{h}(\underset{\sim}{V}) = 0$ yields the L^∞-estimates and Lemma 3.5.16 proves the statement.

□

3.5.4 Graded-channel Approximation

The graded-channel (GC, also laterally asymmetric channel devices) approximation for the computation of currents is an approximation to the current of electrons and holes in a semiconductor where e.g., the electrostatic potential V is already known in a cross section [48, 4, 47, 30]. For a discussion of graded-channel structures, see e.g., [38, 46].

The concentrations of electrons and holes are assumed to be given by Boltzmann distributions

$$p(x) = n_i \exp\left(-\frac{qV(x) - \phi}{k_B T}\right), \tag{3.90}$$

$$n(x) = n_i \exp\left(\frac{qV(x) - \phi}{k_B T}\right) \tag{3.91}$$

where $q > 0$ is the elementary particle charge, $n_i > 0$ the intrinsic charge concentration (for SnO_2: $6 \cdot 10^{17}\text{cm}^{-3}$, see [41]) of the semiconductor, and ϕ the Fermi Level. From the mass action law we obtain that $p(x)n(x) = n_i^2$ [57].

Therefore, by knowing the electrostatic potential V in a cross section A normal to the direction of the charge transport, we are able to find an approximation to the current I.

Considering the drift terms in the drift-diffusion equations we obtain

$$J_n^{\text{drift}} + J_p^{\text{drift}} = -q\mu_n n\nabla V - q\mu_p p\nabla V = qE(\mu_n n + \mu_p p) \tag{3.92}$$

where $E = -\nabla V$ is the electric field.

From this relation we obtain the approximation to the current I via

$$I = \int\int_A J_n^{\text{drift}} + J_p^{\text{drift}}\, dx\, dy = qE \int\int_A \mu_n n(x,y) + \mu_p p(x,y)\, dx\, dy, \tag{3.93}$$

where the x- and y- directions span the cross section A and E is the electric field normal to the cross section, which is assumed to be constant and given by the difference in the applied potentials at the source and drain contacts divided by the length of the structure.

4 Simulations

4.1 Numerical Solution of the Surface Equation

As discussed in Section 2.3, we will consider the three model equations (2.27), (2.30) and (2.24), where we consider equation (2.27) with and without the terms representing the influence of CO. Equation (2.24) provides from a theoretical point of view the most accurate picture. Nevertheless, a lot of work has to be put into the correct estimation of the parameters due to inherent nonlinearity of the functions.

The data used for the estimation were extracted from the measurements that the measurement process is as follows: The current I is held constant at $2\mu A$, the temperature increases at prespecified points and from time to time there is a test gas added (in the underlying data CO). The measurement device then measures the voltage within the nanowire and via Ohm's law ($IR = V$) we obtain the resistance [7, 5].

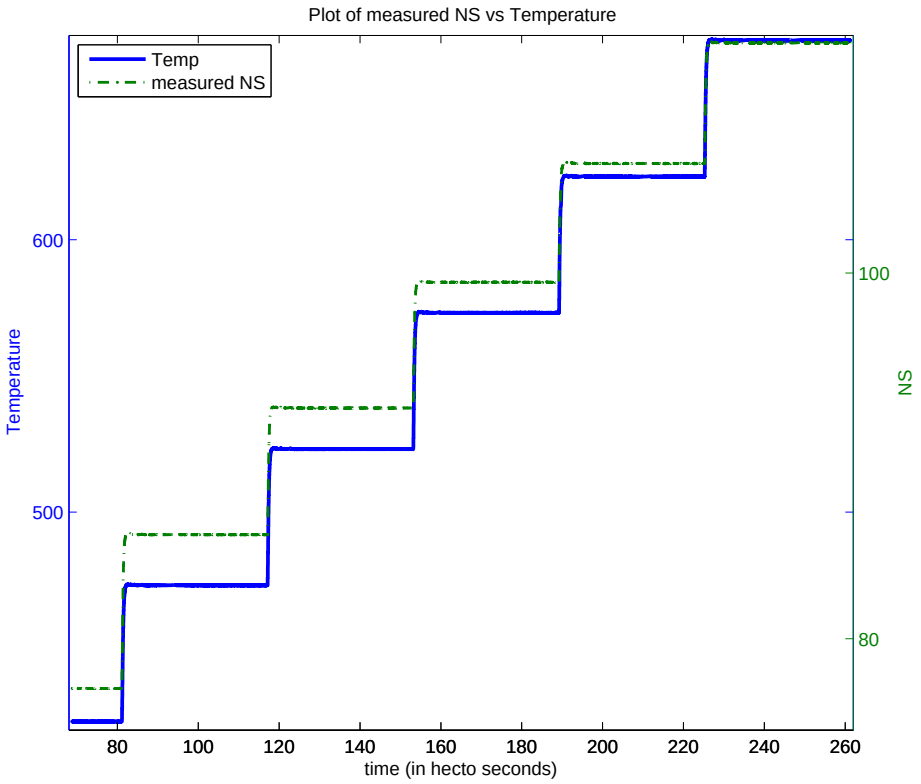

Figure 4.1: Plot of the measured N_S values and the Temperature profile on the surface. As can been derived from equation (2.31) the N_S values increase with increasing temperature.

Equation (2.27) and (2.24) represent assumption A (Section 2.3.4) that the oxygen ionization is the rate-determining step [1]. Equation (2.30) represents assumption B (Section 2.3.4) that the chemisorption is a slow process compared to the ionization reaction [1].

As indicated in Section 2.3.4, we included in all models an additional additive parameter K and denote the models as follows:

1. **FOPNC**:
$$\frac{dN_S}{dt} = A\exp\left(-\frac{N_S^2+\omega_1}{T}\right) - B\exp\left(-\frac{\omega_2}{T}\right)N_S + K, \tag{4.1}$$

2. **FOPC**:
$$\frac{dN_S}{dt} = A\exp\left(-\frac{N_S^2+\omega_1}{T}\right) - B\exp\left(-\frac{\omega_2}{T}\right)N_S - D\exp\left(-\frac{\omega_4}{T}\right)N_S + K, \tag{4.2}$$

3. **FIPC**:
$$\frac{dN_S}{dt} = A\exp\left(-\frac{N_S^2+\omega_1}{T}\right) - B\exp\left(-\frac{N_S^2+\omega_1}{T}\right)N_S - C\exp\left(-\frac{\omega_2}{T}\right)N_S - D\exp\left(-\frac{\omega_4}{T}\right)N_S + K, \tag{4.3}$$

4. **SPC**:

$$\frac{dN_S}{dt} = \exp\left(\frac{-N_S^2}{T}\right) \frac{A\exp(-\frac{\omega_1}{T}) + B\exp(-\frac{\omega_2}{T})N_S}{C\exp(-\frac{\omega_3}{T}) + \exp(-\frac{N_S^2}{T})} - B\exp\left(-\frac{\omega_2}{T}\right) N_S - D\exp\left(-\frac{\omega_4}{T}\right) N_S + K. \tag{4.4}$$

4.1.1 Results of Parameter Identification

The models FOPNC (4.1) and FOPC (4.2) fall short of modelling the main characteristics of the measurement curve. In particular, the computed curve is almost linear which is reflected in the fact that the first term A is the main control factor of the slope of the curve and the other parameters can almost be neglected.

Parameters	**FOPNC**	**FOPC**
A	2.55E+003	9.97E+000
B	5.87E-005	5.54E-005
D		2.90E-010
K	4.79E-003	7.07E-003
ω_1	3.54E+003	1.27E+003
ω_2	3.54E+002	8.35E+000
ω_4		6.40E-002

Table 4.1: The parameter values of the models FOPNC and FOPC.

In particular, it shows that the additional CO term in FOPC has no effect, as the pre-exponential factor is only approximately 10^{-10}.

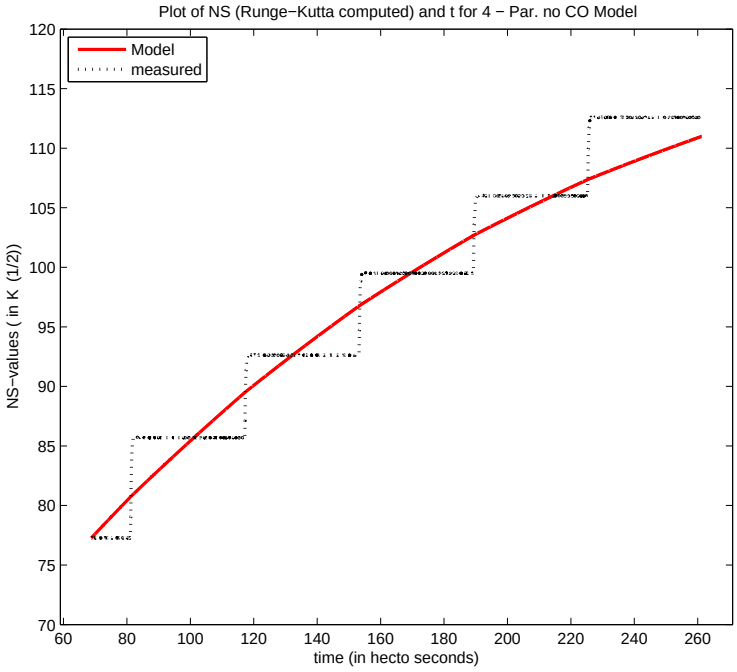

Figure 4.2: Computed values of the FOPNC model (4.1) versus the measured values.

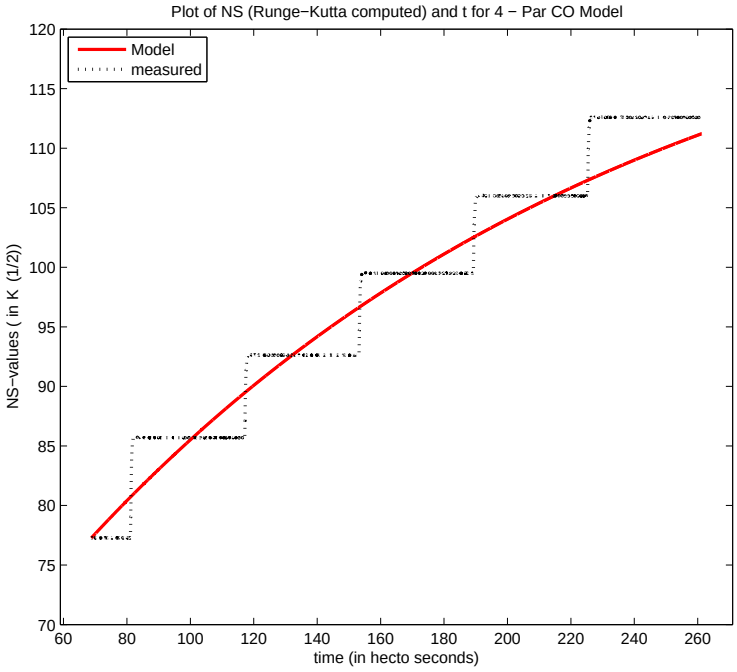

Figure 4.3: Computed values of the FOPC model (4.2) versus the measured values.

The most complex model SPC (4.4) shows its relation to the previous two models (FOPC & FOPNC)[1] by being again mostly dependent on the first factor A. In particular the fractional term has no dramatic effect on the slope of the curve, since ω_3 is very small, and the additional exponential factor in the denominator is small compared to $C\exp(-\omega_3/T)$.

Parameters of SPC	
A	1.00E+000
B	7.21E-013
C	3.01E+001
D	5.56E-005
K	7.17E-003
ω_1	4.22E+000
ω_2	1.73E-003
ω_3	1.18E-008
ω_4	1.55E-003

Table 4.2: The fitted parameter for SPC (4.4).

Furthermore, the values yielded by the model are almost identical to those of the simplified models.

[1]FOPC and FOPNC are derived from SPC. (see Section 2.3.4)

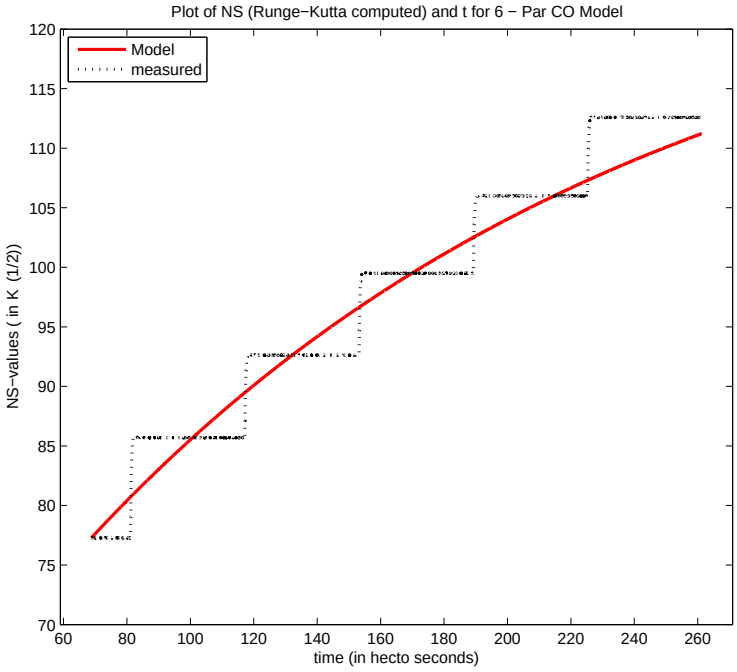

Figure 4.4: Computed values of the SPC model (4.4) versus the measured values.

On the other hand, the FIPC model equation (4.3) provides a very accurate picture of the development of the measured N_S values and, in contrast to the other models, the CO term is of relevance when computing the slope of the function.

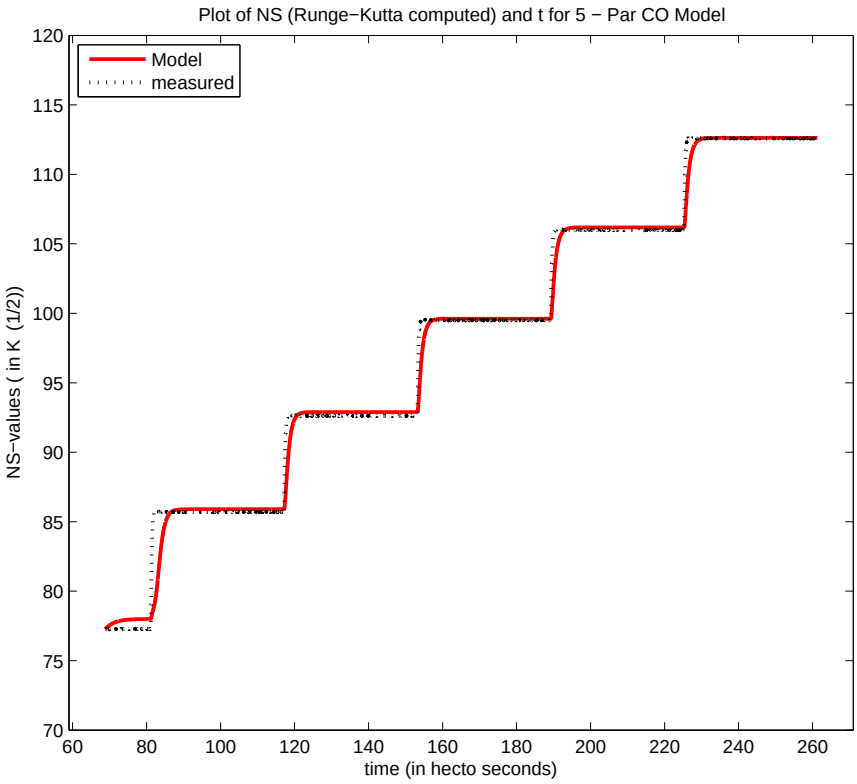

Figure 4.5: Computed values of the FIPC model versus the measured values. Except at the jumps, FIPC describes very accurately the behavior of the measurements.

The computed curve approximates the measured curve after some time adjustment to the temperature within an error range of 0 to 5%. The large errors occurring at the temperature rise cannot be described accurately by just one model for the whole data, because it requires to approximate a step function via a smooth function, which is in general very difficult.

Parameters of FIPC	
A	1.68E+001
B	3.07E+002
C	1.00E+000
D	1.00E+000
K	6.87E-001
ω_1	4.12E-001
ω_2	6.64E+001
ω_4	6.26E+001

Table 4.3: The parameter values for FIPC (4.3).

A further indication of which model represents accurately the underlying measurements can be derived from the difference in size of the parameters. FOPNC, FOPC and SPC show large differences of 10^8 or more within the parameter set, implying inability to represent the characteristics of the measurement curve.

In Table 4.4 we have summarized the results for the estimation of the parameters.

Parameters				
	FOPNC	**FOPC**	**FIPC**	**SPC**
A	2.55E+003	9.97E+000	1.68E+001	1.00E+000
B	5.87E-005	5.54E-005	3.07E+002	7.21E-013
C			1.00E+000	3.01E+001
D		2.90E-010	1.00E+000	5.56E-005
K	4.79E-003	7.07E-003	6.87E-001	7.17E-003
ω_1	3.54E+003	1.27E+003	4.12E-001	4.22E+000
ω_2	3.54E+002	8.35E+000	6.64E+001	1.73E-003
ω_3				1.18E-008
ω_4		6.40E-002	6.26E+001	1.55E-003

Table 4.4: Parameter set for the models FOPNC, FOPC, FIPC and SPC.

4.1.2 Interpretation and Computation of p-Value

When performing statistical tests, one would like to have some kind of measurement value that indicates how strong his or her conviction is about the acceptance or reject of the hypothesis under consideration. For this purpose, statisticians devised the *p-Value* that is the probability of obtaining a test statistic at least as extreme as the one that was actually observed, assuming that the null hypothesis is true [70]. This is in contrast to simple reject or do not reject statements, that do not account for the degree of conviction.

A widely accepted interpretation - many papers published in scientific journals use such an interpretation - is the following:

p-Value	**Interpretation**
$p < 0.01$	very strong evidence against H0
$0.01 \leq p < 0.05$	moderate evidence against H0
$0.05 \leq p < 0.10$	suggestive evidence against H0
$0.10 \leq p$	little or no real evidence against H0

Table 4.5: An interpretation of the p-Value widely accepted in international journals [8].

The smaller the p-Value is, the stronger is the evidence against the null-hypothesis (H0). In the above cases, the null-hypothesis is that the model FIPC is not more accurate than the models FOPNC and FOPC [8].

In order to compute the p-Value, we make use of the F-test. First, we compute the F-ratio via

$$F_r := \frac{(SSA - SSB)/SSB}{(dFA - dFB)/dFB}, \tag{4.5}$$

where SSA is the sum of squares for model A and SSB is the same for model B. dFA is the degrees of freedom for model A and dFB is the same for model B. In our case, SSB represents the sum of squares of FIPC and SSA respectively the other. The degree of freedom dF for the models is computed as follows:

$$dF = \ \#\text{ data points } - \ \#\text{ parameters} \tag{4.6}$$

A basic assumption when applying the F-ratio is that the sum of squares decreases when the number of parameters increases, i.e., the degrees of freedom decrease. That implies that negative F-ratios do not satisfy the basic assumption and the F-test is not applicable [55].

Model	degrees of freedom	sum of squares
FOPNC	23445	11.932690
FOPC	23443	11.932724
FIPC	23442	11.916773
SPC	23441	11.932725

Table 4.6: Degrees of freedom and residual norm of the models.

As can be extracted from Table 4.6, we are able to compute the F-ratios for FIPC versus FOPNC and FIPC versus FOPC. A comparison of FIPC vs. SPC is not possible, because the SOS of SPC is higher than that of FIPC implying that the corresponding F-ratio is negative. The higher SOS given a larger number of parameters is the prime indicator, that the more complex models falls short of reflecting the sensor behavior more accurately than FIPC.

	F-ratio
FIPC vs. FOPNC	10.437357
FIPC vs. FOPC	31.378222

Table 4.7: Table of F-ratios computed for the respective model comparisons.

The corresponding p-Values ($p = 1 - F^{-1}(F_r)$, where F^{-1} is the inverse of the F-distribution) for the model comparisons are given in Table 4.8.

	p-Value
FIPC vs. FOPNC	10^{-6}
FIPC vs. FOPC	$2.147 * 10^{-8}$

Table 4.8: Table of p-Values.

Referring to Table 4.5, we are able to deduce that the assumption that FIPC is not better than FOPC can be dismissed, and the corresponding assumption with respect to FOPNC can be dismissed as well, because the value is below 0.01 implying a very strong evidence against H0, and the equation does not account for the CO influx.

In total, FIPC represents best the measurements, followed by FOPNC, FOPC and SPC if we rank them with respect to their sum of squares (see Table 4.6).

4.2 Numerical Simulations of the Gas Sensors

After constructing a surface model for the surface reactions and the Poisson-Boltzmann model for calculating the current, we now want to relate the results to measurements carried out at the AIT. Our model consists of the following parts:

Given the concentration of electrons on the surface from the surface model, we use the nonlinear Poisson-Boltzmann equation (3.70a) in order to compute the electric potential for a given temperature (in our case 400 degree Celsius).

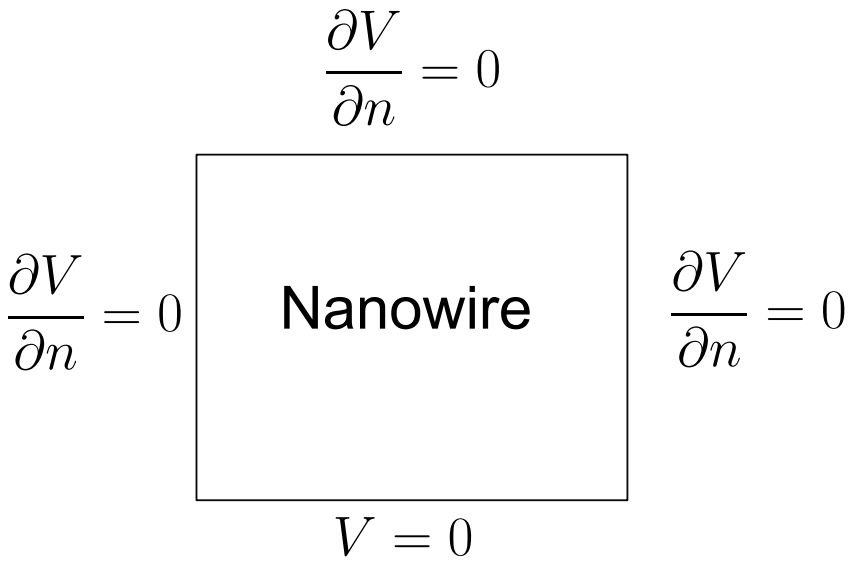

Figure 4.6: Schematic diagram of the simulation domain for the nonlinear Poisson-Boltzmann equation. On three sides we assume zero Neumann boundary conditions and on one side Dirichlet zero boundary conditions.

Discretizing, the solution of the Equation (3.59a) corresponds to finding the zero of the function

$$F(V) := AV + f(V) - C \tag{4.7}$$

where A possesses the pentadiagonal form shown in Figure 4.2,

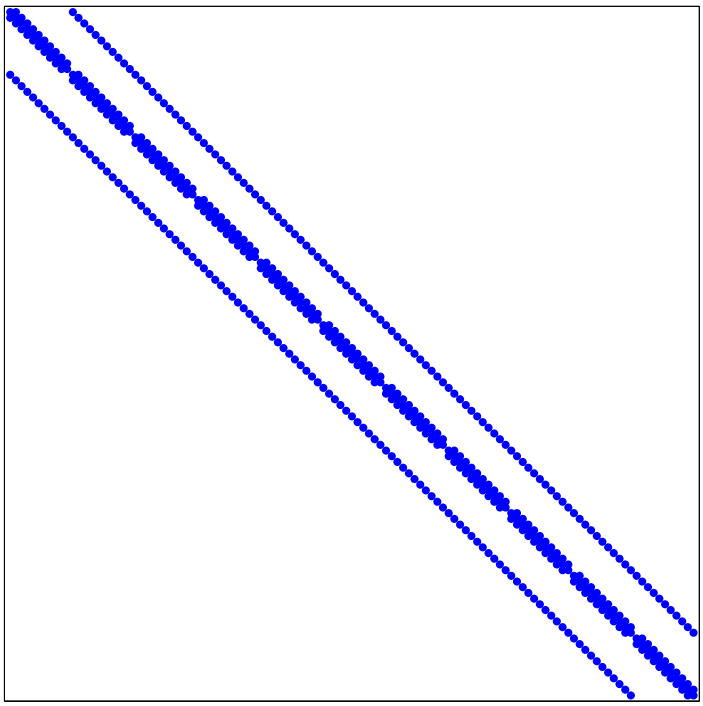

Figure 4.7: Sparsity structure of matrix A upon discretizing the Poisson-Boltzmann equation.

The results of the simulations and their interpretation are given below:

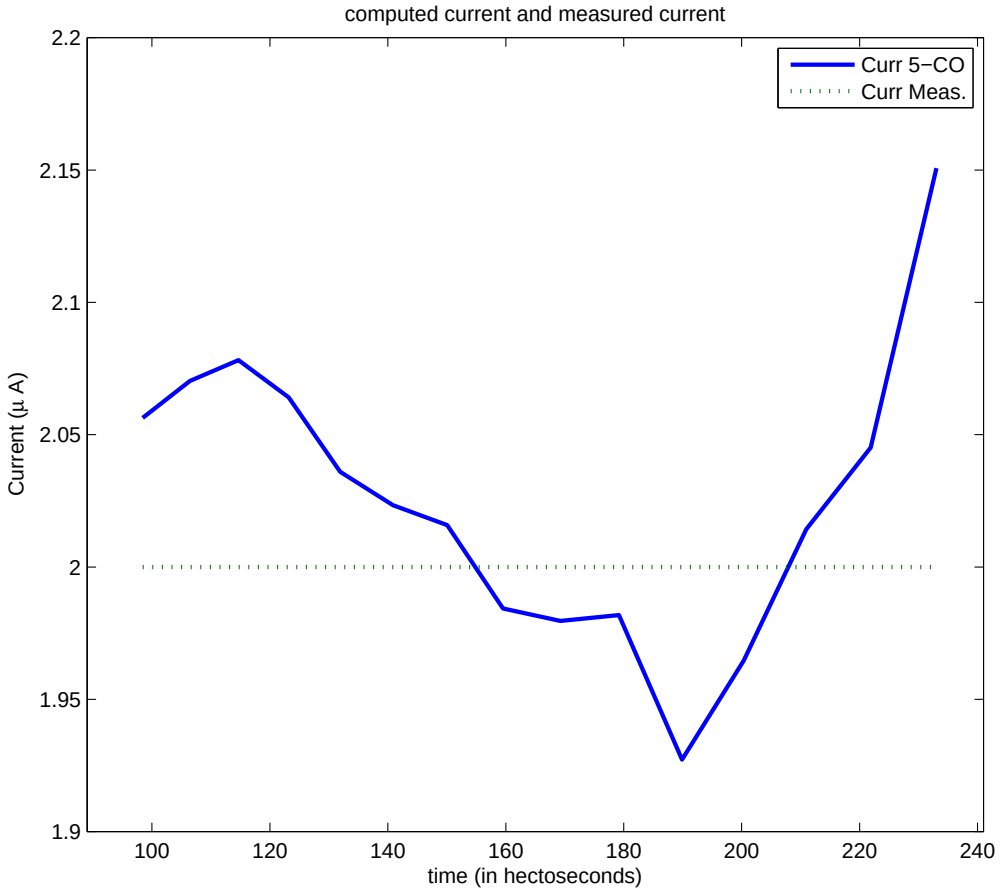

Figure 4.8: Comparison of simulated current to the pre-specified current used for the measurements at AIT.

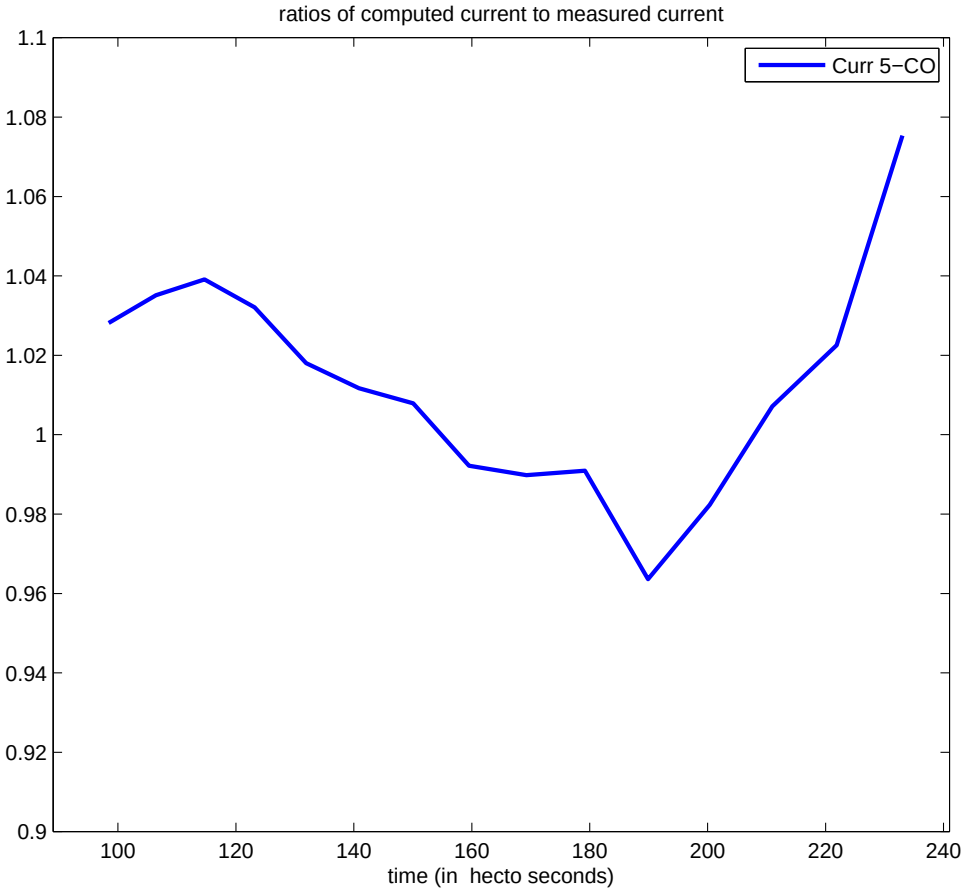

Figure 4.9: Deviation of the simulated current to the pre-specified current.

The corresponding data are given below:

Time (s)	**Voltage** (V)	**Simulated Current, SC (μA)**	**Current, C (μA)**	**Deviation (SC to C) in %**
9841	0.1221	2.0563	2.0000	2.82
10644	0.1230	2.0703	2.0000	3.51
11469	0.1234	2.0782	2.0000	3.91
12317	0.1226	2.0641	2.0000	3.21
13190	0.1209	2.0360	2.0000	1.80
14086	0.1202	2.0234	2.0000	1.17
15007	0.1197	2.0158	2.0000	0.79
15952	0.1179	1.9844	2.0000	-0.78
16926	0.1176	1.9796	2.0000	-1.02
17919	0.1177	1.9818	2.0000	-0.91
18989	0.1145	1.9273	2.0000	-3.64
20035	0.1167	1.9647	2.0000	-1.77
21097	0.1196	2.0143	2.0000	0.72
22186	0.1215	2.0451	2.0000	2.26
23304	0.1277	2.1507	2.0000	7.54

Table 4.9: Data table of the simulated current compared to measured current.

Table 4.2 shows that the simulated current deviates by less than 4% from the measured current for the time between 9.841 and 22.186 seconds and only at the end deviates by a larger factor. Specifically, the simulated current overestimates the measured current up to 15.500 s and then slightly underestimates it up to around 20.500 s.

5 Appendix

5.1 Matlab Code

For this Master thesis were written about 8.000-10.000 lines of code in Matlab. In particular, a least squares optimizer for the parameter estimation and a nonlinear Poisson-Boltzmann solver for the Poisson-Boltzmann model were implemented. The most important files are described below:

lsqOptimizer.m	Least squares optimization routine that is adapted to each equation.
Rk4AIT.m	Runge-Kutta 4 routine for the solution of the reaction kinetics equations.
constrAFNonLinNeuDir2D.m	Sets up the matrix A in equation (4.7) and a vector that contains the other entries not dependent on the potential V
SolvingNonLinPB.m	Solves the zero finding problem $F(V) = 0$ for F as in (4.7).
currentGradChanApprox.m	Computes the current I from the potential provided by SolvingNonLinPB.m

Bibliography

[1] A. Fort, M. Mugnaini, S. Rocchi, S.-S. Rocchi, M. B. Spinicci, R. Vignoli. Surface State Model for Conductance Responses During Thermal-Modulation of SnO2 -Based Thick Film Sensors: Part I – Model Derivation. *IEEE Transactions on Instrumentation and Measurement*, 55(6):2102–2106, 2006.

[2] A. Fort, M. Mugnaini, S. Rocchi, S.-S. Rocchi, M. B. Spinicci, R. Vignoli. Surface State Model for Conductance Responses During Thermal-Modulation of SnO2 -Based Thick Film Sensors: Part II – Experimental Verification. *IEEE Transactions on Instrumentation and Measurement*, 55(6):2107–2117, 2006.

[3] N. Greenwood A. Earnshaw. *Chemistry of the Elements*. Butterworth-Heinemann, 1997.

[4] A. Kranti, T. M. Chung, D. Flandre and J.-P. Raskin. Laterally asymmetric channel engineering in fully depleted double gate SOI MOSFETs for high performance analog applications. *Solid-State Electronics*, 48(6):947–959, 2004.

[5] A. Tischner, A. Köck, T. Maier, C. Edtmaier, C. Gspan, G. Kothleitner. Tin oxide nanocrystalline films and nanowires for gas sensing applications. *Microelectronic Engineering*, 86:1258–1261, 2009.

[6] A. Tischner, A. Köck, T. Maier, M. Kast, C. Edtmaier, C. Gspan, and G. Kothleitner. Atmospheric pressure fabrication of SnO_2-nanowires for highly sensitive CO and CH_4 detection. *Sensors and Actuators B: Chemical*, 138:160–167, 2009.

[7] A. Zima, A. Köck and T. Maier. In- and Sb-doped tin oxide nanocrystalline films for selective gas sensing. *Microelectronic Engineering*, 87:1467–1470, 2010.

[8] H. Arsham. Kuiper's p-value as a measuring tool and decision procedure for the goodness-of-fit test. *Journal of Applied Statistics*, 15(3):131–135, 1988.

[9] N. Barsan and U. Weimar. Conduction model of metal oxide gas sensor. *J. Electroceramics*, 7:143–167, 2001.

[10] N. Barsan and U. Weimar. Understanding the fundamental principles of metal oxide bases gas sensors: The example of CO sensing with SnO_2 sensors in presence of humidity. *J. Phys.: Condensed Matter*, 15:813–839, 2003.

[11] R. G. Bartle and D. R. Sherbert. *Introduction to Real Analysis*. Wiley, 1999.

[12] M. S. Bazaraa. *Nonlinear Programming: Theory and Algorithms*. Wiley, 2006.

[13] A. Bjoerck. *Numerical Methods for Least squares Problems*. SIAM, 1st edition, 1996.

[14] A. Hamnett C. Hamann and W. Vielstich. *Electrochemistry*. Wiley, 1998.

[15] S. L. Campbell and K. D. Yeomans. Behavior of the nonunique terms in general DAE integrators. *Applied Numerical Mathematics*, 28(2-4):209–226, 1998.

[16] F. J. Holler D. A. Skoog, D. M. West and S. R. Crouch. *Fundamentals of analytical chemistry*. Brooks/Cole, 2004.

[17] D. Barrettino, M. Graf, M. Zimmermann, C. Hagleitner, A. Hierlemann and H. Baltes. A Smart Single-Chip Micro-Hotplate-Based Gas Sensor System in CMOS-Technology. *Analog Integrated Circuits and Signal Processing*, 39:275–287, 2004.

[18] N. Davidson. *Statistical Mechanics*. McGraw-Hill, 1962.

[19] L. Debnath and P. Mikusinnski. *Introduction to Hilbert spaces and Applications*. Academic Press, 1998.

[20] J. Dieudonné. *Foundations of modern analysis*. Academic Press, 1969.

[21] E. Comini, G. Faglia, G. Sberveglieria, Z. Pan, Z. L. Wang. Stable and highly sensitive gas sensors based on semiconducting oxide nanobelts. *Applied Physics Letters*, 81(10):1869–1871, 2002.

[22] I. Ekland and R. Temam. *Convex Analysis and Variational Problems*. North-Holland, 1976.

[23] Lawrence C. Evans. *Partial Differential Equations*. American Mathematical Society, 1998.

[24] F. Fogolari, A. Brigo and H. Molinari. The Poisson – Boltzmann equation for biomolecular electrostatics: a tool for structural biology. *J. Mol. Recognit.*, 15(6):377–392, 2002.

[25] R. Fletcher. *Practical methods of Optimization*. Wiley, 1987.

[26] D. Gilbarg and N. S. Trudinger. *Elliptic partial differential equations of second order*. Springer, 2001.

[27] P. E. Gill and W. Murray. Algorithms for the solution of the nonlinear least-squares problem. *SIAM Journal on Numerical Analysis*, 15(5):977–992, 1978.

[28] W. Gopel. Chemisorption and charge transfer at ionic semiconductor surfaces. Implications in designing gas sensors. *Progress in Surface Science*, 20(1):9–103, 1985.

[29] C.M. Guldberg and P. Waage. Concerning chemical affinity. *Erdmann's Journal für Practische Chemie*, 127:69–114, 1879.

[30] H. Kaur, S. Kabra, S. Bindra, S. Haldar and R.S. Gupta. Impact of graded channel (GC) design in fully depleted cylindrical / surrounding gate MOSFET (FD CGT/SGT) for improved short channel immunity and hot carrier reliability. *Solid-State Electronics*, 51(3):398–404, 2007.

[31] B. B. Owen H. S. Harned. *The physical chemistry of electrolyte solutions*. Reinhold, 1958.

[32] F. Holleman and E. Wiberg. *Inorganic Chemistry*. Academic Press, 2001.

[33] M. J. Holst. *Multilevel Methods for the Poisson-Boltzmann Equation*. PhD thesis, University of Illinois at Urbana-Champaign, 1993.

[34] J. Ding, T. J. McAvoy, R.E. Cavicchi, and S. Semancik. Surface state trapping models for SnO_2-based microhotplate sensors. *Sens. Actuators B: Chem.*, 77:597–613, 2001.

[35] J. D. Jackson. *Classical Electrodynamics*. Wiley, 1999.

[36] J.J. More and D.C. Sorensen. Computing a trust region step. *SIAM Journal on Scientific and Statistical Computing*, 3:553–572, 1983.

[37] F. Jones. *Lebesgue Integration on Euclidean spaces*. Jones and Bartlett Publisher, 1993.

[38] K.-H. Su, W.-C. Hsu, P.-J. Hu and Y.-J. Chen. An improved symmetrically-graded doped-channel heterostructure field-effect transistor. *Journal of the Korean Physical Society*, 50(6):1878–1882, 2007.

[39] C. T. Kelley. *Iterative Methods for Linear and Nonlinear Equations*. SIAM, 1995.

[40] C. T. Kelley. *Iterative methods for Optimization*. SIAM, 1999.

[41] C. Koerber. *Herstellung und Charakterisierung polykristalliner kathodenzerstraeubter Zinnoxid-Duennschichten*. PhD thesis, TU Darmstadt, 2010.

[42] G. Korotcenkov. Metal oxides for solid-state gas sensors: What determines our choice? *Materials Science and Engineering: B*, 139(1):1–23, 2007.

[43] E. A. Moore L. Smart. *Solid State Chemistry: An Introduction*. CRC Press, 2005.

[44] K. Levenberg. A method for the solution of certain problems in least squares. *Quart. Appl. Math.*, 2:164–168, 1944.

[45] E.W. Lund. Guldberg and Waage and the Law of Mass Action. *J. Chem. Ed.*, 42:548–550, 1965.

[46] G. Lutz. *Semiconductor Radiation Detectors: Device Physics*. Springer, 1st edition, 2007.

[47] M. A. Pavanello, J. A. Martino, V. Dessard and D. Flandre. Analog performance and application of graded-channel fully depleted SOI MOSFETs. *Solid-State Electronics*, 44(7):1219–1222, 2000.

[48] M. B. Patil, S. N. Mohammad and H. Morkoc. Modeling of field-effect transistors with laterally graded doping. *Solid-State Electronics*, 32(9):791–795, 1989.

[49] M. Nitta, S. Ohtani and M. Haradome. Temperature dependence of resistivities of SnO_2-based gas sensors exposed to CO, H_2, and C_3H_8 gases. *Journal of Electronic Materials*, 9(4):727–743, 1980.

[50] M. Y. Afridi, J. S. Suehle, M. E. Zaghloul, D. W. Berning, A. R. Hefner, R. E. Cavicchi, S. Semancik, C. B. Montgomery, and C. J. Taylor. A Monolithic CMOS Microhotplate-Based Gas Sensor System. *IEEE Sensors Journal*, 2(6):644–655, 2002.

[51] M.A. Branch, T.F. Coleman, and Y. Li. A subspace, interior, and conjugate gradient method for large-scale bound-constrained minimization problems. *SIAM Journal on Scientific Computing*, 21(1):1–23, 1999.

[52] D. Marquardt. An algorithm for least-squares estimation of nonlinear parameters. *SIAM J. Appl. Math.*, 11:431–441, 1963.

[53] D. A. McQuarrie. *Statistical Mechanics*. Harper and Row, 1973.

[54] A. Mordecai. *Nonlinear Programming: Analysis and Methods.* Dover, 2003.

[55] H. Motulsky and A. Christopoulos. *Fitting Models to Biological Data Using Linear and Nonlinear Regression: A Practical Guide to Curve Fitting*. Oxford University Press, 1st edition, 2004.

[56] J. Munkres. *Analysis on Manifolds*. Addison-Wesley, 1991.

[57] D. Neamen. *An Introduction to Semiconductor Devices*. McGraw Hill, 1st edition, 2005.

[58] R. M. Nix. An introduction to surface chemistry. Course notes for a class on Surface Chemistry, available under `http://www.chem.qmul.ac.uk/surfaces/scc/`, 1997.

[59] J. Nocedal and S. J. Wright. *Numerical optimization*. Springer, 1999.

[60] A. K. Pichler-Tennenberg. Sobolev Imbedding Theorems, `http://www-maths.swan.ac.uk/staff/vl/projects/main.pdf`.

[61] R. Precup. *Theorems of Leray-Schauder Type And Applications*. CRC Press, 2002.

[62] M. H. Protter and H.F. Weinberger. *Maximum Principles in Differential Equations*. Springer, 1999.

[63] R. H. Byrd, R. B. Schnabel, and G.A. Shultz. Approximate solution of the trust region problem by minimization over two-dimensional subspaces. *Mathematical Programming*, 40:247–263, 1988.

[64] W. S. Harwood R. H. Petrucci and G. Herring. *General Chemistry: Principles and Modern Applications*. Prentice Hall, 2001.

[65] R. Ionescu, E. Llobet, S. Al-Khalifa, J. W. Gardner, X. Vilanova, J. Brezmes, and X. Correig. Response model for thermally modulated tin oxide-based microhotplate gas sensors. *Sens. Actuators B: Chem.*, 95:203–211, 2003.

[66] W. Rudin. *Principles of Mathematical Analysis*. Mc Graw- Hill, 1976.

[67] S.-J. Chang, T.-J. Hsueh, I.-C. Chen, B.-R. Huang. Highly sensitive ZnO nanowire CO sensors with the adsorption of Au nanoparticles. *Nanotechnology*, 19:1–5, 2008.

[68] S. L. Campbell, C. T. Kelley, and K. D. Yeomans. Consistent initial conditions for unstructured higher index DAEs: A computational study. *Proc. Conference on Computational Engineering in Systems Applications (CESA 96)*, pages 416–421, 1996.

[69] G. Sberveglieri. *Gas Sensors: Principles, Operation and Developments.* Springer, 1st edition, 1992.

[70] Mark J. Schervish. P Values: What They Are and What They Are Not. *The American Statistician*, 50(3):203–206, 1996.

[71] S.H. Hahn, N. Barsan, U. Weimar, S.G. Ejakov., J.H. Visser and R.E. Soltis. CO sensing with SnO_2 thick film sensors: Role of oxygen and water vapour. *Thin Solid Films*, 436(1):17–24, 2003.

[72] H. Kurata T. Kariya. *Generalized Least Squares.* Wiley, 1st edition, 2004.

[73] T. S. Rantala, V. Lantto, and T. T. Rantala. Rate equation simulation of the height of schottky barriers at the surface of oxidic semiconductor. *Sens. Actuators B: Chem.*, 13:234–237, 1993.

[74] C. Tanford. *Physical chemistry of macromolecules.* Wiley, 1961.

[75] U. Pulkkinen, T. T. Rantala, T. S. Rantala, and V. Lantto. Kinetic Monte Carlo simulation of oxygen exchange of SnO_2 surface. *J. Molecular Catalysis A: Chem.*, 166:15–21, 2001.

[76] V. Brynzary, G. Korotchenkov, and S. Dmitirev. Simulation of thin film gas sensors kinetics. *Sens. Actuators B: Chem.*, 61:143–153, 1999.

[77] W. Schmid, N. Barsan, and U. Weimar. Sensing of hydrocarbons and CO in low oxygen conditions with tin dioxide sensors: Possible conversion paths. *Sens. Actuators B: Chem.*, 103(1-2):362–368, 2004.

[78] W. Walter. *Ordinary Differential Equations, (German version).* Springer, 2000.

[79] J. Wolberg. *Data Analysis Using the Method of Least Squares: Extracting the Most Information from Experiments.* Springer, 1st edition, 2005.

[80] A. Zima. *Development of highly sensitive nano-gassensors based on nanocrystalline tin dioxide thin films and single-crystalline tin dioxide nanowires.* PhD thesis, Vienna University of Technology, 2009.

Curriculum Vitae

Name:	Katterbauer Klemens, BSc.
Birthdate:	8.7.1987
Contact:	Margaritenweg. 8
	2301 Gross-Enzersdorf
Phone:	+43 (664) 73 75 48 46
Email:	katterbauer.klemens@gmail.com

Education

03/2010 - present:	**University of Vienna**, Vienna, AUSTRIA ⋆ Mathematics (MSc) *Majors:* Applied Mathematics & Scientific Computing, Optimization
10/2007 - 02/2010:	**University of Vienna**, Vienna, AUSTRIA ⋆ Mathematics (BSc), *Majors:* Applied Mathematics, Optimization finished with distinction ⋆ bachelor papers: *Finite Element Method*, Supervisor: Hermann Schichl *Elliptic Regularity*, Supervisor: Bernhard Lamel
10/2007 - 02/2008:	**University of Vienna**, Vienna, AUSTRIA ⋆ Economics (BSc), not finished
09/2002 - 06/2007:	**business.academy.donaustadt**, Vienna, AUSTRIA ⋆ A-Level, *Major:* Controlling and Accounting finished with distinction ⋆ 2006: Austrian champion in the business simulation game TopSim ⋆ project paper: *Comparison of US-GAAP, IFRS, HGB at Intercell AG.*
04/2006:	**TÜV AUSTRIA** Certified Quality Manager: passed with distinction

SACHERSCHLIESSUNG
der
Fachbereichsbibliothek Mathematik, Statistik, Informatik der Universität Wien

Klassifikation

Aufstellungsort: ☒ HoA ☐ HoD

MSC2000: 92E20 – 80A30 – 80A23 – 90C30 – 90C90 – 65K05 – 65N08 – 35J47

BK: 31.80 – 35.18

Schlagwortketten nach RSWK
Poisson-Gleichung – Methode der kleinsten Quadrate – Grenzflächenchemie
(213, 312)

Kontrollvermerk der Fachbibliothek:

27.10.10
Datum

Unterschrift