

Abstract

Passivated steel can act as fit substrate for thin-film photovoltaics. A thin zirconium oxide layer was applied via blade coating onto a steel foil substrate with the goal of getting a homogeneous and insulating layer. Layers were qualitatively characterized with SEM and XRD and quantitatively characterized via current-voltage curves. The process variables (solution concentration, number of coating layers, coating speed, coating temperature, calcination speed and calcination temperature) were optimized by particle swarm optimization (PSO) algorithm in combination with multivariate adaptive regression splines (MARS). A correlation between the calcination temperature and the electrical properties of the ceramic layers has been revealed. The MARS model performed well compared to linear regression, kernel ridge regression and support vector regression.

Deutschsprachiges Abstract

Passivierter Stahl kann als geeignetes Substrat für Dünnschicht-Photovoltaik fungieren. Eine dünne Schicht aus Zirkoniumoxid wurde mittels Rakelbeschichtung auf ein Stahlfolien-Substrat aufgebracht, mit dem Ziel, eine homogene und isolierende Schicht zu erhalten. Die Schichten wurden qualitativ mit SEM und XRD und quantitativ über Strom-Spannungs-Kurven charakterisiert. Die Prozessvariablen (Precursorlösungskonzentration, Anzahl der Schichten, Beschichtungsgeschwindigkeit, Beschichtungstemperatur, Kalzinationsgeschwindigkeit und Kalzinationstemperatur) wurden mithilfe des Partikelschwarmoptimierungs (PSO)-Algorithmus in Kombination mit multivariaten adaptiven Regression Splines (MARS) optimiert. Eine Korrelation zwischen der Kalzinationstemperatur und den elektrischen Eigenschaften der keramischen Schichten wurde festgestellt. Das MARS-Modell schnitt im Vergleich zu linearer Regression, Kernel-Ridge-Regression und Support Vector Regression gut ab.

Acknowledgment

I thank Alexander for laying a strong material scientific foundation. I thank Leticia, Sebastian and Philipp for introducing me to the intricate world of theoretical chemistry, dynamics and machine learning. I thank Theo, Adhi, Neha, Philipp and Alexander for giving me fruitful advice during my master's project. I thank Quyhn, Maria, Jana, Vivien and Katharina who kept me company in the lab and inspired me. I thank Elif undertook initial experimental part of this project rolling. I thank Karin, Johanna, Dennis and Gala who supported me with finishing this thesis. I thank my parents for my upbringing, teaching me to stay curious and to never give up. Finally, I thank this project which taught me how import is to plan (experiments), chose fitting methods, act reproducibly and how important proper research and constant reading of literature is.

Contents

1	Introduction	8
2	Theoretical Background	9
2.1	Photovoltaics	9
2.1.1	History of Photovoltaics	9
2.1.2	Photovoltaics Basics	9
2.1.3	Copper Indium Gallium Selenide Solar Cell	10
2.2	Sample Preparation	12
2.2.1	Properties of Zirconium Oxide	12
2.2.2	Tape Casting	12
2.2.3	Contact Deposition by Sputtering	12
2.2.4	Scanning Electron Microscopy	12
2.2.5	X-Ray Diffraction	13
2.3	Machine Learning and Statistics	14
2.3.1	Artificial Intelligence and Machine Learning	14
2.3.2	Machine Learning Methods	14
2.3.3	Population Based Optimization Algorithms	17
2.3.4	Evolutionary Model-Based Multiresponse Approach	17
2.3.5	Design of Experiments	18
2.3.6	Analysis of Variance	19
3	Experimental	20
3.1	Substrate Preparation	20
3.2	Production of Precursor Solutions	20
3.2.1	Aqueous Precursor Solutions	20
3.2.2	Buthanolic Solution	21
3.3	Tape Casting	21
3.4	Calcination of ZrO_2	22
3.5	Characterization of ZrO_2 Layers	23
3.6	Data Processing of Electrical Measurements	23
3.7	Pre-Optimization of Coating Procedure	24
4	Computational Details	26
4.1	Evolutionary Model-Based Multiresponse Approach (EMMA)	26
4.2	Fitting Data via Machine Learning	26
5	Results and Discussion	28
5.1	Material Properties of Produced Layers	28
5.1.1	X-Ray Diffractogram of ZrO_2	29
5.1.2	Spectrometry of ZrO_2 Layers	30
5.1.3	Current Voltage Curves of ZrO_2 Layers	30
5.1.4	Pre-Optimization of Coating Procedure	31
5.2	EMMA	31
5.3	Post-EMMA	35
5.3.1	Analysis of Variance	35

5.3.2	Linear Regression	35
5.3.3	Grid Search, Kernel Ridge Regression and Support Vector Machine	36
5.4	Hindsight Is Easier than Foresight	37
5.4.1	Research, Exploration and Design Phase	37
5.4.2	Narrowing, Pre-Optimization and Constraint Phase	38
5.4.3	Optimization, Modeling and Data Generation Phase	38
6	Summary and Outlook	40
A	Pre-EMMA Experiments	48
B	EMMA Experiments	49

Acronyms

- H₂SO₄** sulfuric acid. 12, 21, 28
- N₂** nitrogen. 20
- NaOH** sodium hydroxide. 21, 28
- Zr(PrO)₄** zirconium(IV)propoxide. 20, 21, 28
- ZrO₂** zirconium dioxide. 12, 22–24, 28–31
- 1F** one-fold concentrated solution. 21, 28, 29
- 2F** two-fold concentrated solution. 21
- 3F** three-fold concentrated solution. 21
- 4F** four-fold concentrated solution. 21, 28
- 5F** five-fold concentrated solution. 21
- AcOH** glacial acetic acid. 21, 24, 28, 29
- AI** artificial intelligence. 14
- ANN** artificial neural networks. 15
- ANOVA** Analysis of Variance. 19, 35
- BF** basis function. 34, 35
- BuOH** Butan-1-ol. 20, 21, 28
- CIGS** CuIn_xGa_{1-x}Se₂. 9, 10
- coating** blade coating. 21, 22, 24, 28, 31
- DI water** deionized water. 20, 21
- DOE** design of experiments. 18
- EMMA** evolutionary model-based multiresponse approach. 17, 26–29, 31, 34–38, 40
- EPV** events per predictor variable. 34, 38
- FEG** field emission gun. 12
- FT** Fourier transform. 13
- FTO** fluorine doped tin oxide. 20, 28, 29
- GA** genetic algorithm. 17

GCV generalized cross-validation. 17

HAcAc acetylacetone. 20, 21

HCl hydrochloric acid. 20, 21, 28

I-V current-voltage. 23, 26, 30

IPA Propan-2-ol. 20, 21, 24, 28

ITO indium doped tin oxide. 20

KRR kernel ridge regression. 15, 26, 27, 36

MAE mean absolute error. 15, 27, 36

MARS multivariate adaptive regression splines. 17, 34, 37

ML machine learning. 14, 15, 17

MSE mean squared error. 15, 17, 23, 27, 34–36

NIR near-infrafred. 13, 30

PB Plackett-Burman. 18

PSO particle swarm optimization. 17, 24, 26

PV photovoltaic. 9, 10, 12

RBФ radial basis function. 26

RF regression function. 17, 31, 33–36

SDS sodium dodecyl sulfate. 20, 21, 28

SEM scanning electron microscopy. 12, 13, 20, 28, 30

SVM support vector machine. 15, 16, 26, 27

UV ultraviolet. 13, 30

Vis visible. 13, 30

XRD X-ray diffraction. 13, 29

Chapter 1

Introduction

Photovoltaics (PV) is a viable renewable energy source towards energy neutrality. Furthermore, it uses the sun's energy directly in contrast to other energy sources (e.g. wind, water or even carbon based) and therefore it is fit to be used in energy harvesting projects like futuristic Dyson spheres [1, 2], which harness the whole power output of the sun. One type of thin film PV is based on the CIGS (copper indium gallium sulfide) semiconductor absorber [3]. Due to the large light absorption coefficient of CIGS (compared to silicon), less absorber material is needed (a couple of micrometers) and the PV cell can be made thinner and flexible. In order to make a module, multiple solar cells are connected in series. The solar cells must be deposited on a non-conducting surface. Glass is a popular non-conducting substrate, but rigid and brittle. A flexible alternative is steel foil, which is ductile, inexpensive and highly available, but electrically conducting. An insulating layer must therefore be applied to the steel substrate before any CIGS cells can be deposited on top. Polymers would be a choice if not for their low thermal stability, which does not permit high production temperatures. A non-toxic material which is suitable for the insulation is zirconium oxide (ZrO_2). Sol-gel roll-to-roll coating processes (e.g. tape casting) are economic and scalable methods to apply liquid precursors to substrates, and have been also reported for the deposition of ZrO_2 [4, 5]. Sol-gel processes often produce porous layers, though.

In this work a dense, insulating and homogeneous layer of ZrO_2 is pursued. Machine learning can help to uncover complex non-linear relations, such as the dependence of the thickness and resistance of the deposited layer on the coating parameters. The minimization of the electrical conductance of the coating layer is performed with a particle swarm optimization algorithm which has the coating process variables as input.

This work is organized as follows: chapter 2 gives information on the general working principles of PV, the structure, characteristics and properties of materials which were used and the employed computational, statistical and machine learning methods for data processing.

In chapter 3 and 4 the experimental and computational procedures are described. Section 5 is split into four sections: material specific results are presented in section 5.1, results regarding particle swarm optimization in section 5.2, further analysis in section 5.3 and discussion of the process and the results in 5.4. Finally, chapter 6 summarizes and discusses the outlook and next steps.

Chapter 2

Theoretical Background

This chapter can be broken down into three sections. The first section tries to shine light on the evolution of PV and give some background information on $\text{CuIn}_x\text{Ga}_{1-x}\text{Se}_2$ (CIGS) absorber. The next section concerns the description of materials and scientific methods used during the experimental part of this work. The third and last part focuses on the algorithmic, statistical and analytical methods used to optimize and predict material properties.

2.1 Photovoltaics

The world wide energy consumption has more than doubled between 1970 and 2015 [6] and according to recent studies both fossil [7] and uranium sources [8] will be exhausted within the next 100 years. Even though this time period is not exact and highly dependent on detection methods of resources, the situation demands the fast development of sustainable energy sources. One viable option is photovoltaic (PV).

2.1.1 History of Photovoltaics

The photoelectric effect was first described in 1839 by french scientist Alexandre Edmond Becquerel [9]. Another relevant piece in the PV jigsaw was brought to light with the discovery of photo conductivity of selenium by British engineer Willoughby Smith [10]. In 1876 William Adams and Richard Day [11] showed that the energy of light can be directly converted into electrical energy by a bar of selenium with attached platinum electrodes. And finally, in 1905 Einstein described the physical background of the photoelectric effect with his light quantum theory [12]. In the late 1950s the first solar cells (with power conversion efficiencies around 10 percent [13]) were used in niche applications such as space exploration. In 1958 the US American Vanguard I [14] and the soviet Sputnik III [15] were the first satellites with solar cells. These "solar batteries" - as they were called then - allowed the transmission system of Vanguard I to be operating for over a year after the chemical battery powered system stopped transmitting after 20 days [14]. Eventually, the interest in photovoltaic and other alternative energy sources rose - fuelled by the oil crisis in 1973 - and the development of photovoltaic devices for the consumer market was boosted. This development lead to a drop in average price for PV module from \$ 100 per watt in 1975 to under \$ 1 per watt in 2022 [16].

2.1.2 Photovoltaics Basics

The process of conversion of photons into electric energy can be broken down into two essential steps: the generation of an electron-hole pair through photon absorption, the electron-hole separation by the built in electric field at the p-n junction and finally collection of the electrons and holes at the opposite terminals of the device [17]. This means that PV cells are basically diodes, which have a low resistance in one direction and a high resistance in the other direction. The n-type side of the p-n junction has excess electrons and p-type side has excess electron holes. If the n- and p-type are of the same basis material, they are called homojunctions (e.g. silicon); if not,

heterojunctions (e.g. CdS/CIGS and CdS/CdTe) [18]. The first marketable PV were crystalline silicon photovoltaic modules, which still have the biggest market share in the PV segment (including polycrystalline and monocrystalline silicon). In 2022 over 88% of all sold PV were made of silicon (including amorphous silicon) [18]. A comprehensive overview on different PV technologies can be found in [17].

2.1.3 Copper Indium Gallium Selenide Solar Cell

CIGS ($\text{CuIn}_x\text{Ga}_{1-x}\text{Se}_2$) is of the chalcopyrite (CuFeS_2) group (tetragonal crystal system). CuInSe_2 and CuGaSe_2 were first synthesised in 1953 by Harry Hahn et al. [19]. The potential use of CuInSe_2 as PV material in combination with CdS was first mentioned in 1974 [20]. Cells with efficiencies of over 10% were achieved already by 1975 [21]. Today, CIGS (multiple PV cells in series) reach efficiencies of up to 19%, monocrystalline silicon PV modules reach efficiencies of 24% and polycrystalline silicon PV modules 20% [22].

Just like CdTe, GaAs and amorphous silicon solar cells, CIGS has much higher absorption coefficients of UV, visible and IR light than crystalline silicon (see table 2.1). This is due to a direct band gap rather than an indirect band gap (like crystalline silicon). These thin film PVs not only use less material, but also can be used in flexible applications.

Material	Type	Band Gap [eV]	λ [nm]	Abs. coef. α [cm^{-1}]	Penetration Depth [μm]
c-Si	indirect	1.12	600	4000	2.5
c-Si	indirect	1.12	1000	64	150
c-Si	indirect	1.12	1100	3.5	290
a-Si	direct	1.7	600	40 000	0.25
CdTe	direct	1.45	600	37 000	0.3
GaAs	direct	1.42	600	40 000	0.2

Table 2.1: Photonic properties of several established PV materials (data from [23])

Empirical Formula	Name	Band Gap [eV]	Abbreviation
CuInSe_2	copper indium di selenide	1.04	CISe
CuInS_2	copper indium di sulfide	1.5	CIS
CuGaSe_2	copper gallium di selenide	1.7	CIGSe
CuGaS_2	copper gallium di sulfide	1.55	CIGS

Table 2.2: Band gaps of different chalcopyrites (data from [23])

The band gap of CIGS can be varied between 1 eV and 1.7 eV by varying the indium-gallium and sulfur-selenium ratios. This is a result of the large difference of band gaps of CuInSe_2 and CuGaSe_2 (see table 2.2). In figure 2.1 the schematic layer sequence of a standard CIGS thin film cell is shown. Typically, a 1 μm thick molybdenum layer is deposited on soda lime glass. The sodium in the glass diffuses through the molybdenum layer and increases efficiency and reliability by directing the growth of CIGS in the 112 direction [25]. A CIGS layer of 1-2 μm thickness is applied on top via co-evaporation from elemental sources [25]. The p-type doping of CIGS is achieved by adding more than stoichiometric copper to the mix. The heterojunction is then completed by deposition of CdS n-type layer (typically 50 nm thick). CdS was earlier used as front contact, but now only acts as n-type wide-gap window and buffer. A window layer of intrinsic ZnO is deposited after the CdS layer, followed by a highly conductive aluminium doped layer. ZnO has a band gap of 3.2 eV and is therefore transparent for visible light. The ZnO window layer (usually of thickness 50-70 nm is highly conductive (especially the aluminium doped layer) and acts as front contact.

Instead of glass as substrate, steel and polymers (e.g. polyimide [26]) can be used to create flexible CIGS modules. They both come with their own inconveniences. Polymers generally have low thermal stability compared with ceramics and metals, which restrict the fabrication temperature of CIGS. Steel, on the other hand, is temperature resistant enough, but is an electric conductor,

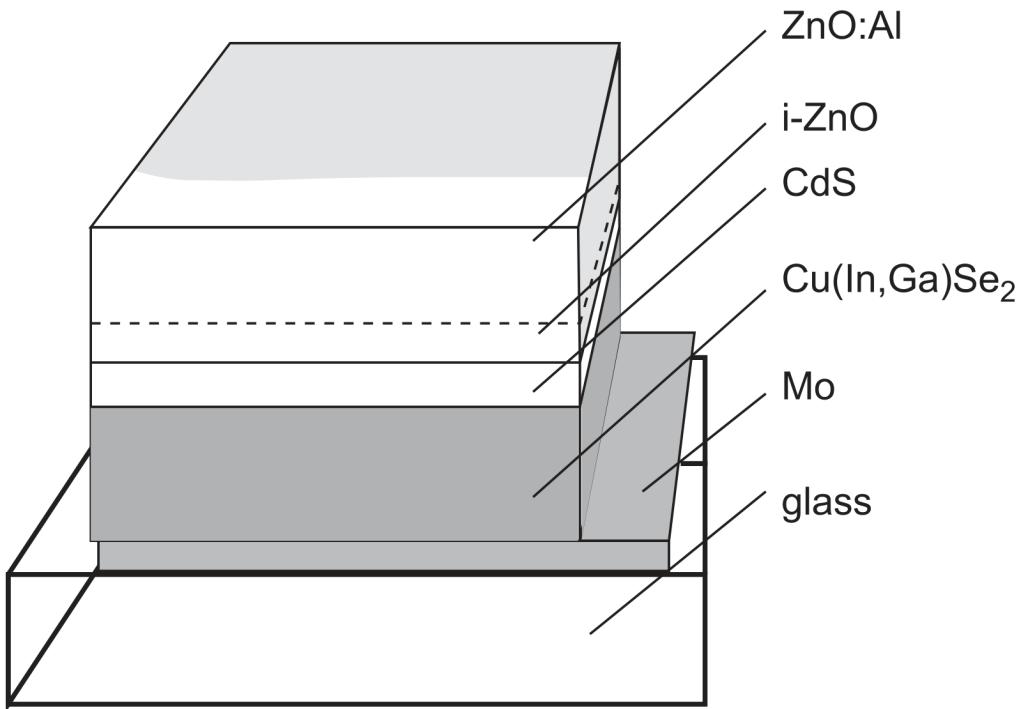


Figure 2.1: Schematic layer sequence of a standard ZnO/CdS/Cu(In,Ga)Se₂ thin-film solar cell [24].

which is problematic for the serial interconnection of cells into modules. Thus, an insulating layer has to be deposited on the steel substrate.

2.2 Sample Preparation

2.2.1 Properties of Zirconium Oxide

Zirconium oxide zirconium dioxide (ZrO_2) is a ceramic with a band gap of 5-7 eV [27] and relative permittivity of circa 20 at room temperature [28]. This makes it attractive as an insulator for semiconductor and PV industry. It is monoclinic below 1050 °C, tetragonal between 1170 °C and 2370 °C, and cubic above 2370 °C [29]. The cubic phase can be stabilized down to room temperature by the addition of magnesia (MgO), calcia (CaO) or yttria (Y_2O_3). This prevents mechanical failing due to shrinkage when cooling and undergoing phase transition [29]. ZrO_2 is very resistant to acids (except HF and hot sulfuric acid (H_2SO_4)) and bases [29].

2.2.2 Tape Casting

Tape casting (also known as tape casting) is widely used in the textile, paper, photographic film, printing and ceramic industries. The roll-to-roll compatible process gives rise to highly uniform films over large areas [30]. A blade is moved over a substrate spreading a slurry at a fixed distance with a fixed speed. In roll-to-roll processes the substrate moves instead of the blade.

2.2.3 Contact Deposition by Sputtering

Sputtering describes the process of highly energetic ions bombarding a surface and atoms being ejected from the surface as a consequence. Use cases vary from thin films depositions for PV, for electrical circuits or for storage media such as CDs and DVDs over sputter cleaning and etching to analysis. Advantages of sputter-deposited thin films include good adhesion to the substrate and good step coverage [31]. The ions which are accelerated to the surface originate mostly from a plasma. The ion bombardment leads to a transfer of momentum from the ions to the surface atoms which in turn leads to a collision cascade. The cascade can reach the surface again. If the energy of an atom is larger than the binding energy, the atom is ejected. This neutral ejected particle travels – unaffected by the electrical field nearly perpendicular to the surface – towards the substrate and condenses with other particles to form a layer. The pressure in the chamber should be small, such that the sputtered particle has a long mean free path, but on the other hand a minimum pressure is needed to sustain the plasma. Usual pressures are around 1 Pa (10^{-2}mbar) or lower [31]. Nobel gases (e.g. argon) in plasma phase are mainly used for bombardment because of their inert properties. Oxygen or nitrogen can be added to the plasma to deposit oxides or nitrides, respectively. In order to use sputtering for depositing a target material on top of a substrate material, the target and the substrate are positioned parallel and in between the gas is transitioned into plasma and the ions are accelerated in the direction of the target (by a high electrostatic potential).

The settings on the Leybold sputtering system were as follows: 40 W for 700 seconds at 0.002 mbar and a argon gas inlet of 6 sccm (standard cubic centimeters per minute).

2.2.4 Scanning Electron Microscopy

Scanning electron microscopy (SEM) is a technique which allows visualization of surfaces with features in the nanometer regime. While optical microscopes use visible light and optical lenses, SEM uses accelerated electron beams and electrostatic and electromagnetic lenses. This allows the generation of much more detailed images due to the shorter wavelengths of electrons compared to photons of visible light [32]. The electron beam produces X-rays, elastically backscattered (primary) electrons, inelastic (secondary) electrons and Auger electrons in the examined material. Secondary electrons carry information to conclude morphology and topology of the sample, while X-rays can be used to identify elements present in the surface. Electrons originate from either a field emission gun (FEG), where a strong electrical field rips electrons from the bulk, or from thermionic guns where a filament (tungsten W or LaB_6 (brighter and longer lasting but more expensive)) is heated until electrons are emitted. Electrons are then accelerated by a voltage of 2 kV to 40 kV and bundled into narrow beams by lenses [33]. All SEM micrographs were taken with a Zeiss Supra 40 with 5 kV acceleration voltages. A high mean free path is needed for electrons

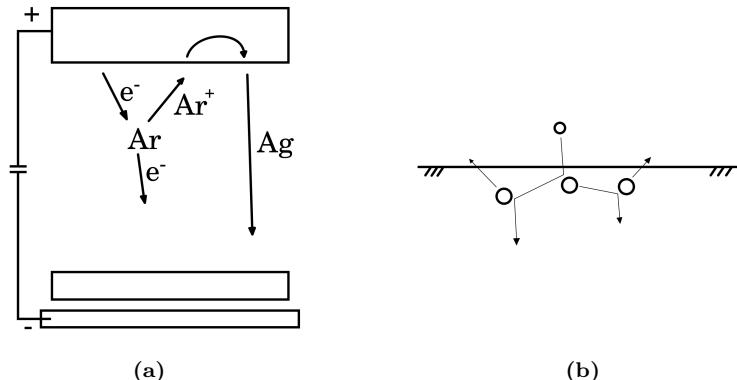


Figure 2.2: Schematic representation of the creation of sputter particles during sputter deposition. (a) creation of ions in the chamber: the top rectangle depicts the source/cathode/target, the bottom rectangle the anode/substrate holder and the middle rectangle the substrate to be sputtered. (b) example series of collision processes leading to sputtering of atoms: horizontal line represents the target surface, the small circle a high energetic ion and the large circles target particles.

to travel from the source to the sample (and to the detector). Thus, a very low pressure is needed inside the microscope. In this work SEM was used as a preliminary way of checking the quality of the deposited layer.

Fourier Transform Spectrometer

In the classic two-beam-spectrometer the light emitted from the light source is split, sent through the sample and the reference, and one of the two beams is alternately sent through a monochromator to a detector (often a thermopile). In the Fourier transform (FT) spectrometer, on the other hand, the beam is sent through the sample, split and reflected from a static and from a moving mirror, recombined and detected by a photo multiplier (a device which transforms photons into electrical signals). The interference of the two beams depends on the optical path difference (also called retardation) of the two light beams [34]. In a FT spectrometer the reference has to be measured before the sample.

FTultraviolet (UV)/visible (Vis)/near-infrared (NIR) transmittance and reflectance spectra (at 20° incident angle) were recorded with a Bruker Vertex 70 spectrometer with a quartz beam splitter, 0.5 mm aperture and Gallium-Phosphide detector for ultra violet light (303 nm–588 nm) and a silicon detector for visual and NIR light (500 nm–1.2 μm). For transmittance the light entered the sample from the side with the coating layer. The UV and Vis/NIR spectra were merged in Opus software.

Reflectance and transmittance were measured at 0° and 20° incident angle, respectively, with a Bruker Vertex 70.

2.2.5 X-Ray Diffraction

X-ray diffraction (XRD) is used to study the crystalline structure of materials. Since X-rays wavelengths (0.2 to 10 nm) are comparable to the interatomic spacing of crystalline solids, the beams get reflected and contain information about the structure [32]. Each crystalline material has a discrete atomic structure which upon irradiation with X-rays causes constructive and destructive interference according to Bragg's law and generates unique diffraction patterns. XRD diffraction plots of crystalline materials feature distinct peaks, whereas amorphous materials exhibit a broad curve with a maximum extending over several degrees (2θ).

XRD diffractograms were obtained with a Thermo Scientific ARL Equinox 100 X-Ray Diffractometer. The diffractograms were measured for 2 min in reflection mode with Cu- K_{α} radiation ($\lambda=1.541\,874\,\text{\AA}$). All XRD diffractograms were taken at 5° incident angle and compared to the internal database.

2.3 Machine Learning and Statistics

Some might argue that machine learning (ML) is just over glorified statistics, but that is not the case [35]. Although ML uses several statistical methods, their goal and frame conditions are different. Statistics and ML are two closely related field and statistics influences ML a lot, but ML is not just statistics with repackages with a new hip buzzword. They differ in their methods, approaches and goals. ML tries to predict unseen data points accurately and statistics is a subfield of maths which tries to get insight into a given data. For example, in a statistical model, it is desirable to reduce the number of inputs. This reduction of inputs allows a statistician to better study how a change in input variables can directly affect an output variable [36]. More precisely, mathematical statistics (stochastics) [37] is the subject of finding mathematical models to describe the data whereas (classical) statistics is the domain of representing the data. Nonetheless, statistics and maths is at the very basis of ML.

2.3.1 Artificial Intelligence and Machine Learning

Artificial intelligence (AI) is a trans-disciplinary field with roots in logic, statistics, cognitive psychology, decision theory, neuroscience, linguistics, cybernetics, and computer engineering [38]. The history of AI goes back to the middle of the 20th century. Researchers from the emerging field came together at a 1956 Dartmouth conference and the term "AI" was coined [39]. AI's history is beautifully depicted by McCorducks' 1982 book "Machines Who Think" [40, 41], which focuses on the great minds behind the advances. Pioneers like Alan Turing thought a lot about how to define, test and implement AI [38]. One example how to measure AI is to let it play chess against a human [42] (in 1997 a chess computer called Deep Blue won against the World Chess Champion Garry Kasparov for the first time [43]). Another test *ingenious* by Alan Turing is the imitation game [44], nowadays known as Turing test: an interrogator communicates with two unknown entities **A** and **B** (a woman and a man) and must find out who is who. **A** will try to make to interrogator misjudge, whereas **B** is on the interrogators side. The question is if **A** is replaced with a computer how the ratio of outcomes would deviate from the original ratio. At the moment it is hard to imagine a computer getting a higher ratio than a human but when reading AI written articles [45], it is easy to see this test being passed in the near future; especially with ChatGPT's convincingly human-like answers [46].

But this does not mean that computers are more intelligent than humans or sentient [47, 48] even though some claim [49] and it certainly does not mean that research is over. AI is still a young field, which is strongly growing and is gaining ubiquitous status. It is slowly creeping into every aspect of modern human life just like electricity around one hundred years ago. Realms in which AI is gaining traction are: playing board games (and beating humans) [42, 43, 50], image recognition (very popular for medical diagnosis) [51–54], chemistry [55–57], cyber security [58], facial recognition (to prevent theft of lavatory paper at public toilet in China [59]), financial sector (as robo-advisors) [60], natural language processing (NLP) [61–63] (which can also create code and pictures through scalable vector graphics (SVG)) and even creative tasks like creating non existing faces [64], create graphic artwork (DALL-E 2) [65] or making video games [66]. It is nearly hard to find a field where AI is not used in some way. This steady incorporation of AI leads to the so called AI effect [40, 67]: certain fields get incorporated into AI research and practice, such that after some time of general use it is no more considered AI (e.g. spam filter or web searches). Google CEO Sundar Pichai even goes as far and said: "AI is one of the most important things humanity is working on. It is more profound than [...] electricity or fire" [68].

2.3.2 Machine Learning Methods

ML is at the base of most AIs. It is an umbrella term for programs with instructions to learn from data, i.e. gain knowledge, categorize, predict and make decisions based on data. There is a platitude of different machine learning methods and most of them can be divided into supervised (training set is labeled) and unsupervised (exploratory). An orthogonal division can be made by regression (continuous data) versus classification (discrete, categorical data). Independent from these 2×2 categories there are multiple ways to let machines learn from data.

Artificial neural networks (ANN) (one of the most popular architectures for big data [69]) are loosely modelled after the brain [70]. Artificial neurons (also called nodes), which are arranged in layers, are connected to each of the neurons of previous and next layers and the weights (parameters of intensity), with which the data is routed from one neuron to another, are optimized during training. Convolutional-layer ANN excel in picture recognition [71] and are useful in quantum mechanics too [72]. Other common methods include linear regression, kernel ridge regression and support vector regression.

Linear regression is one of the simplest methods to predict data. It persuades by its computational simplicity and easy interpretation.

$$\mathbf{y} = \mathbf{X}\mathbf{k} + c \quad (2.1)$$

Where $\mathbf{y} \in \mathbb{R}^n$ (n is the number of data points) is a vector of dependent variables which shall be predicted and $\mathbf{X} \in \mathbb{R}^{n \times d}$ (d is the number of independent variables) is a matrix of independent variables. The parameters $\mathbf{k} \in \mathbb{R}^d$ and $c \in \mathbb{R}$ are chosen by minimizing an objective function (loss function in ML jargon). A typical function to minimize are the L_1 and L_2 error, i.e. mean absolute error (MAE) and mean squared error (MSE), respectively.

$$L_1 = \sum_i |y_i - \hat{y}_i| \quad (2.2)$$

$$L_2 = \sum_i (y_i - \hat{y}_i)^2 \quad (2.3)$$

Ridge regression is like linear regression with an extra term, which penalizes steep regression functions. The extra term reduces overfitting and is scaled by a correction parameter α . The larger α is, the larger is the regularization and the flatter is the regression function. When α tends to infinity, we get an intercept-only model. When α is zero, the regularized loss function becomes the L_2 loss function.

$$L_{RR} = L_2 + \alpha \sum_i |k_i| = \sum_i (y_i - \hat{y}_i)^2 + \alpha \sum_i |k_i| \quad (2.4)$$

Kernel ridge regression (KRR) combines ridge regression with the kernel method. A kernel transforms data in such a way that a linear hyperplane (a point in one dimension, a line in two dimensions, a plane in three dimensions) can fit data in regression problems or separate the data in classification problems without actually doing the transformation for every data point, which lowers computational costs. A kernel is some kind of similarity measure, which fulfills the requirements of non-negativity, symmetry and linearity [73]. The following equations show definitions for linear (eq. 2.5), polynomial (eq. 2.6), sigmoidal (eq. 2.7) and radial basis functional kernel (eq. 2.8), with γ as fixed hyperparameter and c_0 as parameter to optimize. A requirement for using kernels is having a dot product in the loss function. This can be accomplished by expressing \mathbf{k} in terms of \mathbf{X} : $\mathbf{k} = \mathbf{X}^\top \mathbf{r}$ [74].

$$k_{lin}(\mathbf{x}, \mathbf{y}) = \mathbf{x}^\top \mathbf{y} \quad (2.5)$$

$$k_{pol}(\mathbf{x}, \mathbf{y}) = (\gamma \mathbf{x}^\top \mathbf{y} + c_0)^d \quad (2.6)$$

$$k_{sig}(\mathbf{x}, \mathbf{y}) = \tanh(\gamma \mathbf{x}^\top \mathbf{y} + c_0) \quad (2.7)$$

$$k_{rbf}(\mathbf{x}, \mathbf{y}) = \exp(-\gamma \|\mathbf{x} - \mathbf{y}\|^2) \quad (2.8)$$

Support vector machine (SVM) is a versatile machine learning algorithm first mentioned in 1992 [75]. The SVM was initially developed by Vladimir Vapnik for the binary classification of separable data, then improved to handle non-separable data [76] and eventually adapted to solve regression problems [77]. The concepts of SVM will be discussed in the same chronological order. Classification works by spanning a hyperplane between two linearly separable categories in a way such that the closest points from each category have the largest distance to the hyperplane. The distance from the points to the hyperplane is called margin τ . The points with the shortest distance to the hyperplane are called support vectors and are used to define the hyperplane. A SVM avoids overfitting by only using a subset of the data - the support vectors - to fit the model.

The goal is to find the decision boundary which correctly classifies all samples with the biggest margin. The decision boundary can be expressed as a hyperplane

$$\hat{\mathbf{y}} = h(\mathbf{X}) = \mathbf{X} \cdot \mathbf{w} + b \quad (2.9)$$

with $\mathbf{y} \in \{+1, -1\}^n$, $\hat{\mathbf{y}} \in \mathbb{R}^n$, $\mathbf{X} \in \mathbb{R}^{n \times d}$ and $\mathbf{w} \in \mathbb{R}^d$. The constraint of the positive and negative support vectors (\mathbf{x}^+ and \mathbf{x}^- , respectively) satisfying

$$\mathbf{x}^+ \cdot \mathbf{w} + b = 1 \quad (2.10)$$

$$\mathbf{x}^- \cdot \mathbf{w} + b = -1 \quad (2.11)$$

can be generalized to

$$y_i(\mathbf{x}_i \cdot \mathbf{w} + b) \geq 1 \quad (2.12)$$

where y_i being the labels of the training data. The width of the margin can be inferred by projecting the vector spanning between two support vectors on opposite sides of the decision boundary onto the unit vector perpendicular to the hyperplane. If we now take a vector from a positive support vector \mathbf{x}^+ to a negative support vector \mathbf{x}^- and project it onto the unit vector of \mathbf{w} (which is perpendicular to the hyperplane), we get the width of the margin.

$$\begin{aligned} 2\tau &= (\mathbf{x}^+ - \mathbf{x}^-) \cdot \frac{\mathbf{w}}{\|\mathbf{w}\|} = (\mathbf{x}^+ \cdot \mathbf{w} - \mathbf{x}^- \cdot \mathbf{w}) \cdot \frac{1}{\|\mathbf{w}\|} \\ &= ((1 - b) - (-1 - b)) \cdot \frac{1}{\|\mathbf{w}\|} = \frac{2}{\|\mathbf{w}\|} \end{aligned} \quad (2.13)$$

Thus, maximizing the margin is equivalent with minimizing $\|\mathbf{w}\|$ and minimizing $\frac{1}{2}\|\mathbf{w}\|^2$ (mathematical convenience for further steps). By incorporation the constraint (eq. 2.12) via the Lagrangian multiplier method we *mathematically* arrive at the loss function which should be maximized:

$$\mathcal{L} = \frac{1}{2}\|\mathbf{w}\|^2 - \sum_i^n \alpha_i[y_i(\mathbf{w} \cdot \mathbf{x}_i + b) - 1] \quad (2.14)$$

which can be rewritten by some mathematical acrobatics (setting the partial derivatives of the Lagrangian function $\frac{\partial \mathcal{L}}{\partial \mathbf{w}}$ and $\frac{\partial \mathcal{L}}{\partial b}$ to zero and inserting into eq. (2.14)) [78, 79] in the following way:

$$\mathcal{L} = \sum_i^n \alpha_i - \frac{1}{2} \sum_i^n \sum_j^n \alpha_i \alpha_j y_i y_j \mathbf{x}_i \cdot \mathbf{x}_j \quad (2.15)$$

Soft margin SVM – in contrast to hard margin SVM – is used if the data is non-separable due to outliers [76]. Such data can be handled by introducing a penalization term for wrongly categorized samples into the loss function.

$$L_{SM} = \frac{1}{2}\|\mathbf{w}\|^2 + \sum_i^n \max(0, 1 - y_i(\mathbf{w} \cdot \mathbf{x}_i + b)) \quad (2.16)$$

If the prediction ($\mathbf{w} \cdot \mathbf{x}_i + b$) and the true category y_i do not agree, they have opposite signs and their product will be a negative number. The subtraction of a negative number will result in a positive penalization. If the sample is correctly predicted, the product will result in a positive number, the subtraction will lead to a negative number and the maximum will be 0 if the sample is outside of the margin. The function we want to optimize (eq. 2.15) and its soft margin equivalent (eq. 2.16) can both be expressed as inner products. This allows us to separate data not only by hyperplanes but also by intricate decision boundaries due to the kernel trick.

Again, a kernel $K(\mathbf{x}_i, \mathbf{x}_j) = T(\mathbf{x}_i)^\top \cdot T(\mathbf{x}_j)$ allows to calculate the inner product of two vectors in a transformed space without the need of transforming each vector, which turns out to be computationally much cheaper.

SVMs can also be used for regression (support vector regression). The decision function becomes the regression function and the margin includes all data points instead of none. In soft margin support vector regression a few outliers are allowed. Non-linearity can be likewise introduced via the kernel trick.

2.3.3 Population Based Optimization Algorithms

There are also lesser known ML algorithms such as evolutionary algorithms (e.g. genetic algorithm (GA) and particle swarm optimization (PSO)). Evolutionary algorithms take advantage of the ability to cope with local optima by evolving several candidate solutions simultaneously [80]. One particular feature of evolutionary algorithms is that they start with a small data set and periodically request new data in order to solve the problem iteratively.

A GA is a search algorithm that uses principles of natural selection and genetics (mutations and recombination) to optimize a search space. A GA starts with a population of randomly generated solutions, or chromosomes, and then proceeds to breed them together to create new solutions. The new solutions are then tested for fitness, and the best solutions are selected to create the next generation of chromosomes. This process is repeated until a satisfactory solution is found (mostly after a predefined number of repetitions).¹

A PSO also uses a starting population of particles where each experiment (particle) is represented by its independent and dependent variables. It was originally inspired by the behavior of bird flocks and fish schools [80, 81]. Each particle has an associated position and speed. Every movement across the search space is, additionally to a stochastic term, influenced by its particle speed and position as well as its and the swarm's best visited position.

2.3.4 Evolutionary Model-Based Multiresponse Approach

Evolutionary model-based multiresponse approach (EMMA) is an implementation of PSO in the R programming language. Each time step the dependent variables for all possible input variable combinations are predicted with the help of multivariate adaptive regression splines (MARS). The MARS regression function (RF) is then used to chose the next position for each particle. MARS is a regression method introduced by Friedman's 1991 paper "Multivariate Adaptive Regression Splines" [82]. The CRAN (Comprehensive R Archive Network) package which implements MARS is called `earth` due to MARS being trade marked [83]. MARS is an extension of multivariate linear regression. Friedman presented MARS as an alternative to piecewise polynomials (splines) and local averaging methods (e.g. kernel functions). Advantages of MARS are that it uses less parameters and it produces continuous models with continuous derivatives. Furthermore, MARS has the ability to fit (possibly complex) interactions and non-linearities without losing interpretability.

The RF is developed in a forward/backward stepwise recursive manner [84]. During the forward recursion, terms are added in pairs until a certain number of terms is reached. Each pair consists of two hinge functions (also called rectifier functions) multiplied with a term already part of the RF (including the constant term). The maximum degree of interaction is 2 for EMMA, but can be varied. This means that only two basis functions can be multiplied (excluding the constant term) to form a subsequent term. Pairwise added hinge functions are of the simple form $h(x - c)$ and $h(c - x)$ where $h(e) = \max(0, e)$ and with x being an independent variable, c being a constant and e being any expression. Then, the backwards algorithm regularizes the function by removing individual terms. The metric which decides if a term should be removed is called generalized cross-validation (GCV) and was introduced by Wahba and Craven in 1969 [85]:

$$GCV(M) = \frac{1}{N} \sum_{i=1}^N \frac{(y_i - \hat{f}(x_i))^2}{\left(1 - \frac{C(M)}{N}\right)^2} \quad (2.17)$$

with M being the number of terms, N the number of data points, a correction term

$$C(M) = (d + 1)M + 1 \quad (2.18)$$

and penalty is $d = 3$ for interactions larger than 1 and $d = 2$ otherwise. The variable of the hinge functions and its knot location at the forward step and which terms to delete at the backwards step are selected by minimizing GCV. The coefficients for each term are then chosen via regular MSE minimization [84].

¹This paragraph was written by GPT3 [62] given the input "Introduction to genetic algorithms:"

2.3.5 Design of Experiments

”The real purpose of experiment design is to maximize the information content of the data within the limits imposed by the given constraints.” - Grahem C. Goodwin [86]

In two cases a deliberate design of experiments (DOE) is especially beneficial. (1) If the query of a new data point is very expensive, it is favourable to actively chose the query (e.g. drilling for oil or quantum chemical calculations).

(2) If the query space is so vast, that randomly querying might explore domains, which might lead to uninteresting or even misleading information. At the beginning of any experiment its constraints must be determined. Constraints for a given experiment include range of input and output variables as well as total time available and total number of samples/experiments that can be taken [86].

A naive approach to an experiment design is the full factorial DOE. Each possible combination of discrete values is tested. The full factorial design will mostly be infeasible due the curse of dimensionality [79] despite it being the most informative design. The 2-level factorial design provides an alternative with 2^d experiments (where d is the number of independent variables). Drawbacks of 2-level factorial DOEs include no data about the inside of the search space and infeasibility for high dimensional problems. In a full factorial or 2-level factorial design most experiments are redundant and most resources will be spent exploring high-order interaction effects [87], which are often minimal to non-existent. In order to overcome these obstacles a certain number of experiments can be chosen randomly from the search space. When a subset is chosen from the factorial design, it is called a fractional factorial design. The Plackett-Burman (PB) [88–90] design is a special case of 2-level fractional factorial design, where the number of needed experiments n is $n < d + 4$ (more precisely $n = (\lfloor d \div 4 \rfloor + 1) \cdot 4$, where $\lfloor x \rfloor$ denotes the floor function on x). The PB design ensures that each combination of levels for any pair of factors appears the same number of times. A drawback of 2-level factorial (incl. PB) and random fractional designs is that the sample set is likely not evenly distributed across the search space [91]. The Hammersley design [91, 92] is based on the Hammersley sequence and produces space filling data points. The Latin hypercube DOE [91, 92] is a type of orthogonal DOE, which has the feature that each level for each variable will be tested only once. A Latin hypercube DOE can be also created such that data points distribute more uniformly over the search space. Latin hypercube DOEs are mainly used in computer simulations, which are purely deterministic and therefore are very precise.

2.3.6 Analysis of Variance

Analysis of Variance (ANOVA) is a statistical test for estimating the influence of multiple categorical independent variables on a dependent numerical variable. At the heart of ANOVA lies the F-test. The F-test was *ingeniered* by Ronald Fisher [93], an important figure in modern statistics. The F-test uses a ratio of variances to determine if a null-hypothesis (observed difference is due to chance alone) is true. The variance in the numerator measures the "between-group variability" and the denominator measures the "within-group variability". This ratio is unaffected by units, scaling errors and constant bias. Additional assumptions of ANOVA include: groups and levels should be independent, residual error should follow normal distribution and variance within groups should be equal (homoscedasticity).

Chapter 3

Experimental

In this section the used chemicals, substrates, experimental procedures and any used lab equipment are described. All chemicals used in this work are of reagent-grade purity and without further purification. First, the base recipe and the base process settings were explored. Then, the analysis of samples is discussed and the boundaries for the optimization were investigated during pre-optimization. Finally, the experiments for the optimization were performed.

3.1 Substrate Preparation

Five different substrates were used throughout this work: 1 mm-thick, soda-lime microscope glass slides ($2.5\text{ cm} \times 7.5\text{ cm}$) from Sigma Aldrich, thinner, squared glass plates ($2.5\text{ cm} \times 2.5\text{ cm}$) Menzel Gläser, indium doped tin oxide (ITO) coated glass ($2.5\text{ cm} \times 2.5\text{ cm}$) from Sigma Aldrich with a sheet resistance of approximately $12\text{ }\Omega/\text{sq}$, fluorine doped tin oxide (FTO) coated glass ($5\text{ cm} \times 5\text{ cm}$) from Sigma Aldrich and steel foil ($10\text{ cm} \times 10\text{ cm}$) provided by Sunplugged GmbH (<http://sunplugged.at/>). The steel foil was cut with a foil cutter into $2.5\text{ cm} \times 2.5\text{ cm}$ squares. Glass was used because of its low price point and availability and to measure transmission and reflectance spectra. FTO and ITO coated conducting substrates were used for SEM investigation and for electrical measurements (serving as one of the electrical contacts). The conducting substrate allows for easy measurement of electric conductance in vertical direction. Furthermore, while taking a SEM micrograph the sample would quickly accumulate negative charge and repel further electrons without a conducting substrate. All substrates were cleaned in three steps before usage:

1. 15 min in 50 mL deionized water (DI water) and 1 mL of Hellmanex III alkaline concentrate in an ultrasonic bath
2. 15 min in DI water in an ultrasonic bath
3. 15 min in Propan-2-ol (IPA) in an ultrasonic bath

After the last cleaning step, the samples were blown dry with dry nitrogen (N_2) gas and kept in a clean plastic container until the coating step (cf. [94, 95]).

3.2 Production of Precursor Solutions

All recipes for solutions can be divided into two categories: the first recipe - adopted from Anwar et al. [27] - was based on zirconium(IV)propoxide ($\text{Zr}(\text{PrO})_4$) in acetylacetone (HAcAc) and DI water (aqueous solution). The second recipe - adopted from Hu et al. [96] - was based on $\text{Zr}(\text{PrO})_4$ in Butan-1-ol (BuOH) (buthanolic solution).

3.2.1 Aqueous Precursor Solutions

$\text{Zr}(\text{PrO})_4$ was added to acetylacetone (HAcAc) while stirring and in a separate vessel DI water (including any optional additives such as sodium dodecyl sulfate (SDS), hydrochloric acid (HCl),

H_2SO_4 or sodium hydroxide (NaOH)) was added to isopropyl alcohol (IPA) and both were stirred for one hour. These additives were added to influence pH and surface tension with the goal to influence the homogeneity of the resulting layer. The DI water-IPA mixture was added to the $\text{Zr}(\text{PrO})_4$ -HAcAc solution and stirred over night. The exact volumes can be taken from table 3.1. Unfortunately, this solution failed to produce homogeneous coating layers, but instead only resulted in sparse patches of deposits on the substrate. Thus, an alternative solution was found.

Table 3.1: Compositions of different aqueous solutions

recipe	1	2	3	4	5	6	7
$\text{Zr}(\text{PrO})_4$ [mL]	8	8	8	8	8	8	8
HAcAc [mL]	8	8	8	8	8	8	8
IPA [mL]	2	2	2	2	2	2	2
DI water [mL]	2.6	2.6	2.5	2	2	2	2
SDS [mg]	-	5.9	-	-	-	-	-
HCl [mL]	-	-	-	-	0.5	-	-
H_2SO_4 [mL]	-	-	-	-	-	0.5	-
NaOH [mL]	-	-	-	-	-	-	0.5

3.2.2 Buthanolic Solution

Five different concentrations of the buthanolic solutions were prepared. The base recipe (one-fold concentrated solution (1F)) was inspired by Hu et al. [96]. BuOH, $\text{Zr}(\text{PrO})_4$, HAcAc and IPA/glacial acetic acid (AcOH) were used as solvent, precursor, chelating agent and stabilization agent, respectively.

The other four buthanolic solutions (two-fold concentrated solution (2F), three-fold concentrated solution (3F), four-fold concentrated solution (4F) and five-fold concentrated solution (5F)) were similar but with higher concentrations of $\text{Zr}(\text{PrO})_4$ (see table 3.2) with the aim of producing thicker coating layers. Solvents (butane-1,2-diol, BuOH and Propan-1-ol) and chelating agents (HAcAc and citric acid) were tested and later in the process - just before starting the optimization of the process variables - the stabilization compound (AcOH) was changed to IPA. The most promising combination of solvent, chelating agent and stabilization reagent was BuOH, HAcAc and IPA, respectively, which was used for the final optimization.

Table 3.2: Composition of different buthanolic solutions

recipe	1F	2F	3F	4F	5F
BuOH [mL]	4.95	4.9	4.85	4.8	4.75
$\text{Zr}(\text{PrO})_4$ [mL]	0.05	0.1	0.15	0.2	0.25
HAcAc [mL]	0.0125	0.025	0.0375	0.05	0.0625
IPA/AcOH [mL]	2	2	2	2	2

The solvent (BuOH) was put into a beaker glass (or similar, preferably with an air-tight cap) with a magnetic stirring bar and $\text{Zr}(\text{PrO})_4$ was added while stirring. After stirring 15 min one mole equivalent chelating agent (HAcAc) was added and stirred for another 15 min. Finally, the stabilization solvent [96] (IPA or AcOH) was added to the mixture and stirred for additional 30 min. In order to make a 2F solution, the volume of $\text{Zr}(\text{PrO})_4$ and HAcAc was doubled and the volume of BuOH was decreased by the increase of volume of $\text{Zr}(\text{PrO})_4$.

3.3 Tape Casting

All glass substrates were tape-casted manually with a smooth stainless steel wire bar coater. On two opposing edges adhesive tape was applied. This allowed material to only be deposited on the area between the tape. After the coating layer was applied and dried, the tape was removed and the substrate heat treated (see table 3.3). Lower blade coating (coating) velocities resulted in



Figure 3.1: (a) Temperature regulator (1) on the left, vacuum pump (2) in the background and Erichsen Coatmaster 510 film applicator (3) with heatable vacuum plate. (b) Close up of the coating blade in position with blade height adjusted to 0.2 mm. The majority of the suction areas is sealed with tape to increase the underpressure at the remaining ones.

less homogeneous coating layers. Thus, the coating speed was not altered in the manual coating process. One and two layers of adhesive tape were used to alter the depth of the valley.

Steel substrates were doctor bladed with an Erichsen Coatmaster 510 film applicator with a heatable plate. All experiments were performed with a blade height of 0.2 mm. The heatable vacuum plate had equally spaced circular 2.5 mm diameter patches of porous metal where under-pressure could be applied to keep the substrate in place. Most of them were covered with tape (see figure 3.1) to increase the suction intensity at the remaining ones. After setting the temperature of the heating plate to 200 °C, the temperature of the vacuum plate (room temperature, 40 °C, 50 °C, 60 °C, 70 °C or 80 °C) and the coating speed (10 mm/s, 12 mm/s, 14 mm/s, 16 mm/s, 18 mm/s or 20 mm/s), the blade was put into its initial position, the sample was placed on the vacuum plate and the vacuum was switched on. During pre-optimization, lower coating velocities (0.1 mm/s, 0.5 mm/s, 1 mm/s, 2 mm/s and 5 mm/s) were tested. 100 µL of solution were applied with a 10 - 1000 µL pipette and the blade moved over the sample distributing the liquid evenly. After evaporation of the solution, the vacuum was turned off, the pushing bar was put into initial position, the blade was removed and excess solution was removed from the plate with a wipe. The coated substrate was dried for 5 min on a 200 °C heating plate. This process of applying a ZrO₂ layer was repeated as many times as needed.

3.4 Calcination of ZrO₂

A LabTech EH45C heating plate and a Naberterm LB410 muffle furnace were used to calcinate the coated samples. The heating plate could hold temperature for a certain amount of time, but heated with a fixed rate of 10 °C/min. In order to achieve a lower overall heating rate several temperature ramps and plateaus were alternated (see table 3.3a). This heating procedure was called HP1. The HP1 procedure was optimized for the available hardware by a colleague working on the project prior to the author.

The NT1 heating program was used to mimic the HP1 heating procedure from the heating plate in the Naberterm muffle furnace. NT2 is a simplification of NT1 and programs NT3 - NT6 are the same as NT2 with altered heating rate and NT5 additionally used a calcination temperature T_{Cal} of 500 °C. NT2-NT6 had 2 variables (heating rate and one calcination temperature) in contrast to NT1 which had 4 (three different heating rates and calcination temperature). All heating programs were held at the calcination temperature for one hour.

Table 3.3: Heating programs

(a) Heating program HP1 used with the Labtech EH45C heating plate

T [°C]	80	100	150	160	170	180	190	200	250	300	350	400
t [min]	10	10	5	5	5	5	5	10	10	10	10	60

(b) Heating programs NT1 - NT6 used with the Naberterm LB410 muffle furnace

Name	80-150°C [°C/min]	150-200°C [°C/min]	200°C-T _{Cal} [°C/min]	T _{Cal} [°C]	t _{Cal} [min]
NT1	2	1	2	400	60
NT2	2	2	2	400	60
NT3	3	3	3	400	60
NT4	4	4	4	400	60
NT5	4	4	4	500	60
NT6	1	1	1	400	60

3.5 Characterization of ZrO₂ Layers

All characterization procedures were executed at room temperature.

The current-voltage (I-V) curves were measured with Agilent 4156C Precision Semiconductor Parameter Analyzer from -0.5 V to 0.5 V with steps of 10 mV.

Top contacts for the I-V measurements were formed with a Leybold UNIVEX450C Sputter System. by sputtering 200 nm of aluminum through a shadow mask, defining circular contacts of a diameter approximately 6 mm. The bottom contacts to the steel substrate or to the ITO- or FTO-coated glass were formed by removing the ZrO₂ deposit from a corner of the substrate, by soft abrasion with sand paper, and then applying conductive silver paste. See figure 3.2c for a sketch of the connectivity.

The measured I-V curves therefore represent the vertical resistance of the deposited layer. This resistance will be predominantly influence by the presence of shunts (hotspots or short-cuts between the top and bottom contact), appearing due to the inhomogeneous thickness, and pinholes through the ZrO₂ layer.

3.6 Data Processing of Electrical Measurements

Multiple I-V curves were measured (see figure 5.4) for every sample. The main difficulty faced in processing the data was to present the measurements obtained from a sample in one representative number. For every I-V curve (i.e. for every aluminium dot contact) the gradient g at $V = 0$ (conductance) is calculated by taking 5 points before and 5 points after the origin with minimum one point distance from zero (which boils down to the data points from ± 0.02 V to ± 0.07 V), calculating averages \bar{V} and \bar{I} of each point cloud and taking the ratio of differences:

$$g = \frac{\bar{I}_+ - \bar{I}_-}{\bar{V}_+ - \bar{V}_-} \quad (3.1)$$

The average of each conductance would be an easy choice but difficult to represent a sample correctly since the possible values of conductance span across several magnitudes. So the average of base 10 logarithm of conductances is the next nearby *ansatz*. In order to make the metric closer MSE (and to penalize deviation from ideal more than linearly) the average of the squared difference to the ideal non-conduction case ($g = 10^{-13}$) was chosen instead of the average of logarithms and shall be called conductivity hereafter.

$$\gamma = \sum_i^N \frac{(-\log_{10}(g_i) - 13)^2}{N} \quad (3.2)$$

Another measure is the density of shorted elements ρ (pinhole ratio), calculated in following way:

$$s_i = \begin{cases} 1 & \text{if } -\log_{10}(g_i) < 5 \\ 0 & \text{if } -\log_{10}(g_i) \geq 5 \end{cases} \quad (3.3)$$

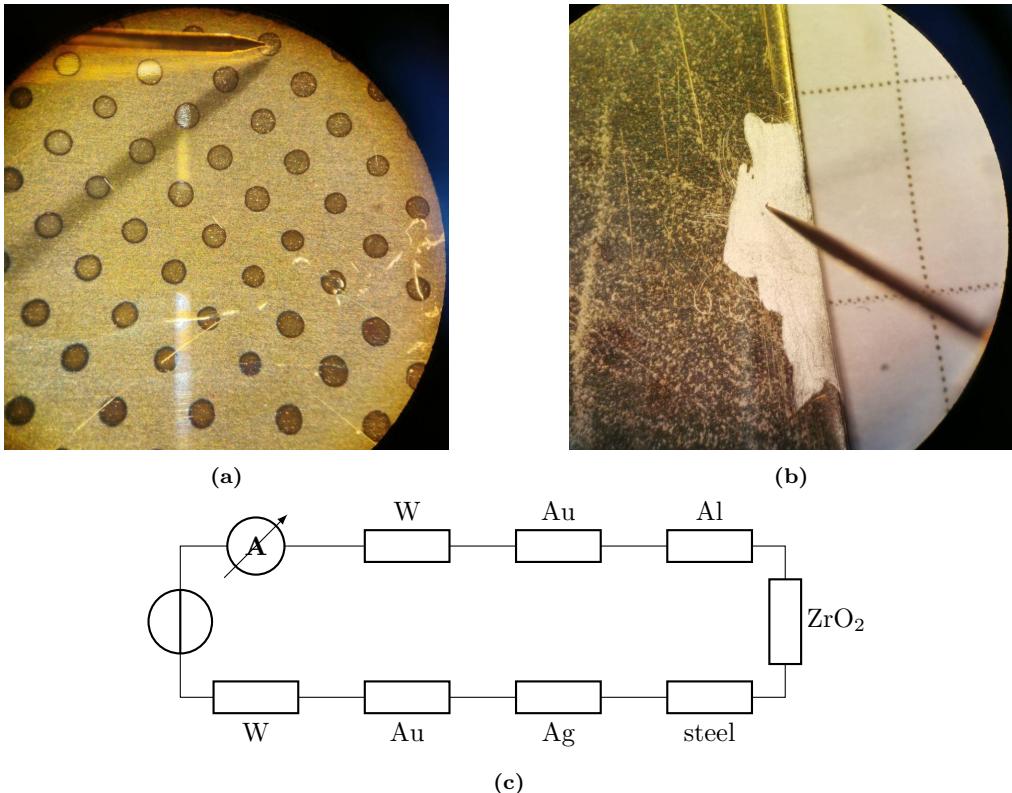


Figure 3.2: View of a sample through the microscope attached to the multimeter and sketch of circuit.
 (a) Sputtered silver contacts on sample with gilded wolfram contact probe touching one of the contacts.
 (b) Edge of a sample with the probe touching silver paste on top of a sanded patch.
 (c) From the voltage source clockwise: variable voltage source, built-in amperometer, gold plated tungsten probe, sputtered aluminium contact, ZrO_2 layer to be measured, steel substrate, dried silver paste contact, gold plated tungsten probe.

$$\rho = \sum_i^N \frac{s_i}{N} \quad (3.4)$$

The density is 1 if every measurement has a conductance larger than $10^{-5} \Omega^{-1}$ and is 0 if every measured conductance has a value of smaller than $10^{-5} \Omega^{-1}$.

3.7 Pre-Optimization of Coating Procedure

A metal plate with vacuum chuck and temperature control was used as a base in the film applicator. The bounds of the process variables were then explored in a preliminary study using the Plackett-Burman [97] design implemented in the python3 (version 3.9) library pyDOE (version 0.3.8); with (1,5), (4,10), (400,500), (120,480), (0.1,5) and (20,80) as nominal and extreme values for relative concentration of Zr c_{zr} , number of coating layers n_L , calcination temperature $T_{cal}[\text{°C}]$, heating rate $v_{cal}[\text{°C/h}]$, coating speed $v_C[\text{mm/s}]$ and coating temperature $T_C[\text{°C}]$, respectively. After testing the first samples, the lower limit for coating speed was altered from 0.1 to 1 (see appendix A).

Some additional hand picked experiments (see appendix A) were introduced to further narrow down the limits of the optimization as every reduction in variables or levels meant a faster convergence. The stabilization agent was changed from AcOH to IPA shortly before the PSO because of much better solution stability and similar electric properties of resulting layers (see samples 192, 199 and 201 in appendix A). The best process variables up to this point were chosen to be tested with 1 mL AcOH (sample no. 199), 0.5 mL AcOH plus 0.5 mL IPA (sample no. 201) and 1 mL IPA (sample no. 192). The sample produced with IPA as stabilization agent showed comparable electric insulation results (see appendix A) and better stability of the solution. Therefore, IPA was used

as stabilizer in further experiments. After the base recipe for the solution has been decided the process variables to produce an insulating layer was examined. See table 4.1 for the input space spanned during the optimization.

Chapter 4

Computational Details

4.1 Evolutionary Model-Based Multiresponse Approach (EMMA)

An evolutionary approach was chosen, namely a multi-objective PSO algorithm called EMMA [80, 81, 98, 99]. Initially, the input parameters (independent variables), their boundaries and number of equidistant levels for each parameter are declared (see table 4.1). Next, the output variables (dependant variables), their weights in the objective function (the function which should be optimized) are specified and if they should be minimized or maximized is noted.

c_{zr} [22 mmol/L]	n_L	T_C [°C]	v_C [mm/s]	T_{cal} [°C]	v_{cal} [°C/h]
2	4	40	10	300	120
3	6	50	12	400	360
4	8	60	14	500	600
5	10	70	16		840
	12	80	18		1080
				20	

Table 4.1: Discrete levels of each input parameter

The *optimizands* where chosen to be γ , ρ , n_L and v_{cal} , (i.e. conductivity, pinhole ratio, number of coated layers and heating rate) where γ , ρ and n_L are to be minimized and v_{cal} to be maximized. The next step is to generate an initial population (ensemble of experiments), which is chosen randomly from the population space. The samples are produced, the I-V curves are measured and the measurements evaluated according to section 3, yielding the quantities γ , ρ_s , n_L and v_{cal} , which were used as inputs for the program.

The program, then, uses this data to estimate a response for each output variable for the entire population space (and to choose a fraction of the initial population which is allowed to propagate). The current population - each of the particles independently - are given a random velocity vector associated to them, which is directed to the best predicted sample. The population for the next time step is generated and again the samples are produced, the I-V curves are measured and measurements evaluated. Each time step thus includes experiment identification, solution preparation, coating deposition and coating layer analysis.

4.2 Fitting Data via Machine Learning

Python 3.10.9 and sci-kit learn 1.1.2 [100] were used to implement a linear fit model, KRR and SVM with polynomial, radial basis function (RBF) and sigmoid kernels.

A grid search over a set of hyperparameters was executed. The hyperparameters which were examined in KRR were kernel, γ_{ml} and α (regularization parameter). The kernel took the form of RBF, sigmoid or polynomial with degree 1–3. The parameters α and γ_{ml} (subscript to distinguish

from *optimizand*) took the values 0.01, 0.02, 0.05, 0.1, 0.2, 0.5, 1, 2, 5 and 10. Note that γ_{ml} is used differently in each kernel.

The kernels for SVM were the same as for KRR. SVM uses C instead of α as regularization parameter) and ε as SVM tube size parameter. C and ε took the values 0.05–1 with step size 0.05. SVM additionally used the **scale** for γ_{ml} . **scale** equals to $\frac{1}{d \cdot \sigma(\mathbf{X})}$ with d being the number of features and $\sigma(\mathbf{X})$ the variance of the input variables.

The input variables ($c_{zr}, n_L, v_C, T_C, v_{cal}, T_{cal}$) of the EMMA data set were scaled with the routine `sklearn.preprocessing.StandardScaler().fit_transform()`. For each calculation the 5-fold cross-validation score was calculated. The score equals to the coefficient of determination R^2 , which is defined the following for non-linear regression functions: ¹

$$R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)}{\sum(y_i - \bar{y}_i)} \quad (4.1)$$

For each method (KRR and SVM) and each *optimizand* (γ and ρ) the parameters with the highest score were noted.

Finally, the different data sets were predicted for each model and MAE and MSE were calculated. These values were also calculated for EMMA and the linear model in order to compare the different methods.

¹see https://scikit-learn.org/stable/modules/generated/sklearn.kernel_ridge.KernelRidge.html

Chapter 5

Results and Discussion

In this section the results will be presented, explained and discussed. Starting with physical measurements and material properties. Followed by the presentation of the results obtained by optimization and exploration with the EMMA method. Finally, alternative regression methods were applied to the data and the results are compared with the results of the EMMA method.

5.1 Material Properties of Produced Layers

As already indicated in the experimental section the aqueous recipe adopted from Anwar et al. [27] failed to produce a homogeneous layer of ZrO_2 . The aqueous solution started out as opaque slurry in contrast to the buthanolic recipe which became opaque only after degrading. Furthermore, the resulting layers were patchy and laced with cracks (see figure 5.1a). Altering the ratio of ingredients, adjusting pH (by adding NaOH , H_2SO_4 or HCl) or adding SDS as surfactant did not increase the measured resistance through the resulting ZrO_2 layers. Even after the consecutive application of two coating layers, there was no significant improvement and the SEM images (see figure 5.1a) showed highly inhomogeneous deposition, with deposited areas next to other areas where the conductive substrate (ITO or FTO) can be clearly seen uncoated. Figure 5.1a shows a typical detail view of ZrO_2 (dark) on top of FTO (finely polycrystalline). Large cracks, non-coated areas and inhomogeneity were shared characteristics among all samples created with the aqueous recipe.

The sol-gel recipe reported by Hu et al. [96] and further optimized, led to improved results (see section 3.2.2). Figure 5.1b shows a top view of 10 coating layers produced from the buthanolic solution on top of a steel substrate. The large irregularity on the bottom left could depict a pinhole, a hole in the coating layer that reaches to the substrate. These irregularities are rather the exception. Alternatively, the small, hardly visible cracks spread across the surface could depict such pinholes.

As soon as the solution showed initial cloudiness, it was declared as unstable and used no more for coating. The solution becoming opaque indicated build-up of precipitation and was not further used. The visually asserted stability of the solution could be increased a lot by replacing the stabilizing agent AcOH with IPA. A 1F and a 4F solution with AcOH - sealed with Parafilm - was stable for approximately 24 h and 2 h, respectively. Whereas a 4F solution with IPA as stabilizing agent stayed stable for circa 96 h. An increase of the $\text{Zr}(\text{PrO})_4$ concentration accelerates the aging process of the solution. The aging process was accelerated by absence of proper sealing. The aging process depends on the water in the air corresponding to an hydrolyzation process [96]. The introduction of IPA into the recipe allowed higher concentrated solutions to be practicable as they could be stored until the next day, instead of producing a new, separate solution for each day. It was found that after a 4F AcOH solution has shown precipitation (became opaque) the addition 50% of the volume in IPA can even reverse the aging effect for a respectable amount of time. This effect is not only due to dilution as additional solvent (BuOH) did not re-stabilize the solution, even after 5 fold dilution.

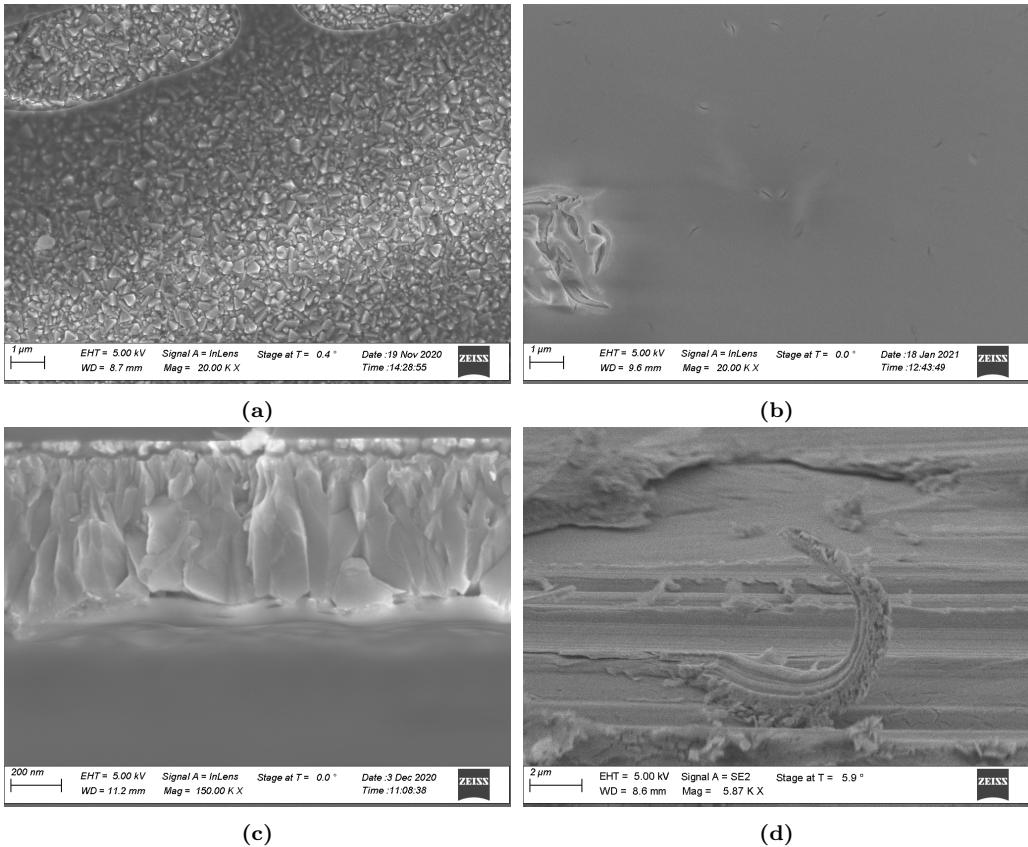


Figure 5.1: SEM images of ZrO_2 : (a) ZrO_2 produced by 2 layers aqueous solution on FTO (b) 10 layers of ZrO_2 on steel substrate (c) cross-section of crack of 5 layers of ZrO_2 on FTO (d) side view of scratched ZrO_2 on steel substrate

Figure 5.1c shows the cross-section of a layer of ZrO_2 on FTO glass at a deliberate fracture. The large crystalline structure, which makes up most of the upper part of the image is FTO. The boundary to glass is visible at the very top. On the lower edge of FTO a circa 100 nm thick homogeneous layer of ZrO_2 (produced with 5 layers of 1F buthanolic AcOH recipe) can be observed. Measuring the cross-section on steel was not as straight forward. Scratching the surface with a diamond glass-cutter allowed us to get an insight into the thickness of the layer on steel substrate. The result can be seen in figure 5.1d. 5 layers of double concentrated solution were applied to this sample. Assuming that the raised structure resembling a tentacle is ZrO_2 , the thickness of the layer can be estimated to be in the order of magnitude of 100 nm.

5.1.1 X-Ray Diffractogram of ZrO_2

XRD was used in order to confirm that the produced layer is indeed ZrO_2 . Figure 5.2 shows diffractograms of the substrate (steel) and the substrate with ten layers of double concentrated solution (EMMA experiment number 6113, see Appendix B). In addition, two idealized diffractograms from the Crystallography Open Database of cubic ZrO_2 (COD ID 1521753 [101]) and of α -Fe (COD ID 1100108) are depicted. The four largest peaks of the cubic ZrO_2 reference diffractogram at $2\theta = 30^\circ, 51^\circ, 60^\circ, 35^\circ$ (highest to lowest intensity) can also be observed in the experimental diffractogram. They confirm that the coated material is indeed ZrO_2 . Raman spectra are needed to further clarify the phase, since cubic and tetragonal phases are difficult to distinguish from XRD alone [102].

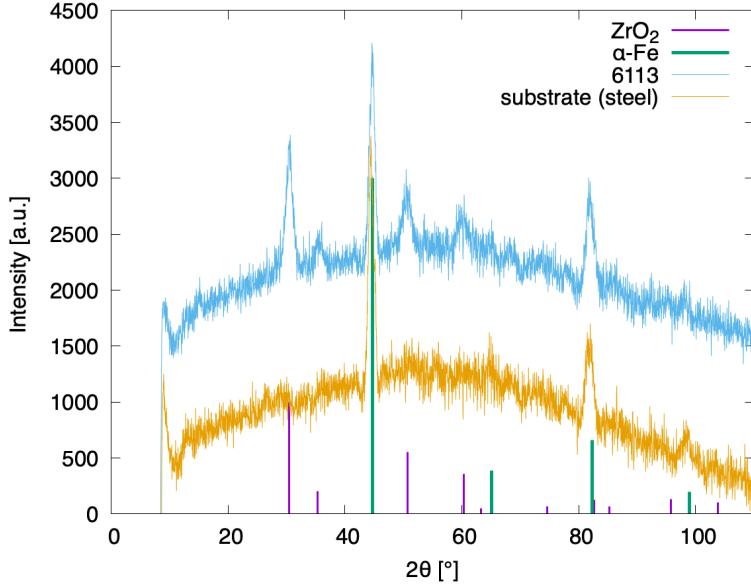


Figure 5.2: XRD spectra of steel substrate (yellow; below), ZrO_2 (blue; above), $\alpha\text{-Fe}$ (green; wide; idealized) and cubic- ZrO_2 (purple; narrow; idealized) [101]

5.1.2 Spectrometry of ZrO_2 Layers

In figure 5.3a UV/Vis/NIR transmittance spectra of ZrO_2 layers on glass slides can be seen. Each sample had different numbers of coated layers. The incident angle was 0° for each sample. The more coating layers, the more of the light is absorbed (at 600-1100 nm) by the ZrO_2 layers. The thicker the layer is, the more turning points can be identified on the graph, which can be attributed to interference [103]. This trend can also be observed in figure 5.3b. The weak interference patterns at low wavelengths indicate that the thickness of the film is in the order of magnitude of the wavelengths [104] – agreeing with SEM measurements. Figure 5.3b shows reflectance spectra of UV/Vis/NIR at an incident angle of 45° . Thicker coating layers reflect more light between approximately 800 nm and 1000 nm. Again, this effect is most likely not observed across the whole spectrum because of interferences.

5.1.3 Current Voltage Curves of ZrO_2 Layers

Figure 5.4a shows I-V curves of an insulating sample, where the abscissa shows the voltage in Volt and the ordinate shows the current in Ampere. Each line represents a I-V measurement at a distinct aluminium contact applied by sputtering through a mask (see figure 3.2). All of the curves show a max voltage of under 10^{-6} V and around a fifth exhibit hardly any current, which is ideal (see figure 5.4d). A single example for a very low conductance measurement in non-logarithmic scale can be seen in figure 5.4g. The conductance (i.e. the gradient at $V=0$) was calculated (see section 3.6) for each measurement. Figure 5.4d shows the distribution of gradients g for the same sample as depicted in 5.4a. Figures 5.4b and 5.4e show measurements of a moderately insulating sample. Most of the I-V curves have a maximum voltage of under 10^{-6} V, but there are some pinholes with conductance above the threshold of 10^{-5} V. In figures 5.4a and 5.4b the minimum of some curves is not at 0 V, which means that the function does not cross the origin in non-logarithmic representation. This deviation (see figure 5.4h) looks similar to the deviation due to the photo-currents [105] because of its non-zero current at $V = 0$. Though, the explanation by photo-current is compelling, it does most likely not apply because of ZrO_2 's band gap of 5 eV [106]. A better explanation would be the tunneling effect, whose I-V curves look very similar [107, 108]. Finally, figures 5.4c and 5.4f show a sample where all measurements exhibit relatively high voltages and high calculated conductances. This indicates a overall bad condition of the ZrO_2 layer for insulating. Figure 5.4i shows a I-V curve which can be approximated by Ohm's law.

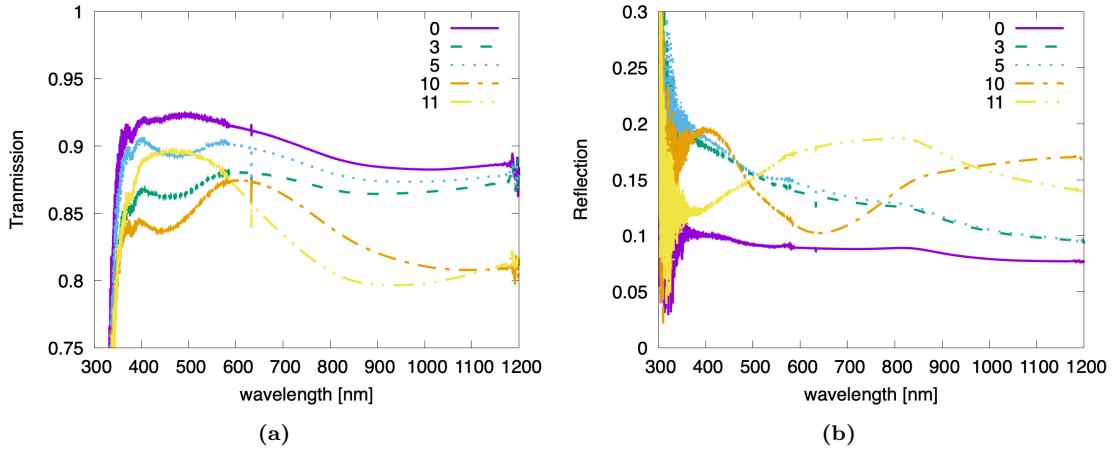


Figure 5.3: UV/Vis/NIR spectra of different ZrO_2 layer counts on glass (a) transmission at 0° and (b) reflection at 45°

5.1.4 Pre-Optimization of Coating Procedure

Before the optimization with the EMMA algorithm, the boundaries of the input variables were explored. Especially, the coating speed v_C (i.e. the speed with which the blade moves and spreads the solution over the sample) was varied and examined during pre-optimization. The slower the coating speed, the more uniformly the solution evaporated. If the v_C was too slow (less than 1 mm/s), no ZrO_2 layer was formed. The reason could be that the force exerted by the blade on the liquid is not strong enough to overcome the combination of gravity and the surface tension of the alkoxide solution. This means that behind the blade a meniscus would pull the liquid without leaving behind any solution for the gelling process. Additionally, the coating temperature affects the evaporation process together with the coating speed v_C .

5.2 EMMA

A total of 30 recipes (see appendix B) was investigated in five iterations ($t = 0, \dots, 4$) of the algorithm. Where the first generation encompassed 10 particles and each subsequent generation encompassed 5 particles. The recipes that produced the most insulating coating for each generation can be seen in table 5.1. The experiments for generation 5 where not executed but predictions were already made with the information from the previous generations. The sample predicted as optimum by the algorithm is experiment number 13 with lowest possible c_{zr} , v_C , T_C , v_{cal} , T_{cal} and second highest possible n_L (see table 4.1).

Table 5.1: Global optimum per generation with experiment numbers and experiment conditions

generation	measurements	enr	c_{zr}	n_L	v_C	T_C	v_{cal}	T_{cal}
1	10	1	2	4	10	40	120	300
2	15	5	2	6	10	40	120	300
3	20	2947	4	6	16	80	1080	300
4	25	2405	2	6	10	40	1080	300
5	30	13	2	10	10	40	120	300

The only clear trend from table 5.1 is the count of ZrO_2 layers count, which rises with the generations of the EMMA algorithm. The remaining input variables remained more or less the same except for the 3rd generation. It can be seen in table 5.2 that the predicted conductivity $\hat{\gamma}$ (predicted with 5th generation RF; see last column) lowered with each iteration, except for sample 2947, which was not predicted but measured. This shows that the easiest part of the algorithm - the selection of the optimum from predicted values - works as expected.

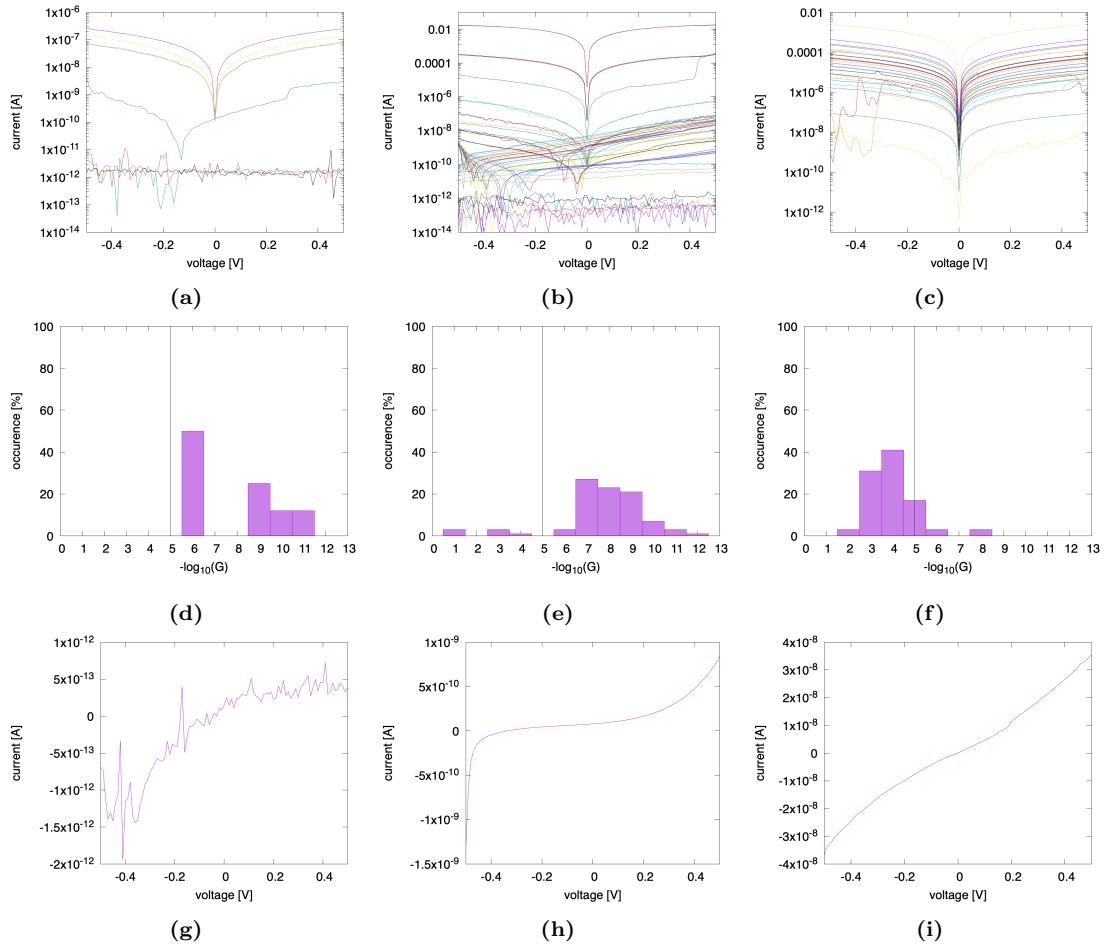


Figure 5.4: Current-voltage curves and the distribution of gradients for three representative samples: insulating well (a),(d),(g), moderately (b),(e),(h), poorly (c),(f),(i); (a)-(c) I-V curves with log transformed current; (d)-(f) frequency histograms of the conductance with pinhole threshold marked as vertical line; (g)-(i) individual I-V measurements

Table 5.2: Predicted conductivity measure $\hat{\gamma}$ for each generation by each generation's regression function

enr	1st gen RF	2nd gen RF	3rd gen RF	4th gen RF	5th gen RF
1	1.214185				38.7962
5		4.196626			25.47335
2947			10.9594		10.9594
2405				20.04962	25.47335
13					24.87178

Contrarily, the predicted conductivity $\hat{\gamma}$ for each generation's optimum (predicted by the very generation's RF) does increase with each iteration (see diagonal in table 5.2). This indicates an underestimation of $\hat{\gamma}$ at the beginning and a correction with time. The underestimation probably stems from a lucky selection of initial experiments or a skew in the measurements. Indeed, the samples with the lowest *optimizands* are among the initial generation (see figure 5.5a, figure 5.5b and appendix B).

In figures 5.5a and 5.5b we can see the two measured main *optimizands*, conductivity γ and pinhole ratio ρ (see section 3.6), of each particle at each generation. The solid circles at generation number 0 indicate particles which were not included to be propagated. Solid lines connect individual particles in time. The dashed line connects averages of particles of each generation (including initial non-propagated particles) and pointed lines connect averages of each generation of only those particles which were propagated. Both γ and ρ were to be minimized and show a clear trend towards low values with increasing generation, indicating that the optimization worked even though the prediction functions (see equations (5.1)–(5.8)) and the chosen samples (i.e. chosen input variables) were not exactly as expected. Neither expected were the measurements for these samples. The deviation from expectation might be due to measuring error of samples and variation of quality due to uncontrolled independent variables such as room temperature, humidity or solution age. Measurements were expected to show clear correlation of *optimizands* γ and ρ with mainly c_{Zr} and n_L . The only statistically relevant correlation was observed for T_{cal} with the *optimizands*. In the scatter plots 5.5c and 5.5d the correlation of calcination temperature T_{cal} with γ and ρ can be seen clearly.

Equations (5.1)–(5.4) and (5.5)–(5.8) represent the regression functions at $t = 3$ and $t = 4$, respectively, rounded to 2 significant digits. The expression $h(n_L - 6)$ translates into number of coating layers only has an influence if larger than 6 and $h(6 - n_L)$ into number of coating layers is influential only if under 6. The first thing to notice is that prediction functions of each generation depend on the same variables. This stems from the fact that the algorithm chooses a single minimal set of basis functions to predict all dependent variables.

$$\hat{\rho}_3 = 0.075 - 0.0014 \cdot v_{cal} + 0.18 \cdot h(6 - n_L) + 3.9 \cdot 10^{-6} \cdot v_{cal} \cdot T_{cal} \quad (5.1)$$

$$\hat{\gamma}_3 = 43 - 0.097 \cdot v_{cal} + 10 \cdot h(6 - n_L) + 0.00026 \cdot v_{cal} \cdot T_{cal} \quad (5.2)$$

$$\hat{n}_{L3} = 9.9 - 0.00064 \cdot v_{cal} - 2.7 \cdot h(6 - n_L) - 1.3 \cdot 10^{-6} \cdot v_{cal} \cdot T_{cal} \quad (5.3)$$

$$\hat{v}_{cal,3} = -5.2 \cdot 10^{-15} + 0.016 \cdot v_{cal} + 1.3 \cdot 10^{-15} \cdot h(6 - n_L) + 3.9 \cdot 10^{-21} \cdot v_{cal} \cdot T_{cal} \quad (5.4)$$

$$\hat{\rho}_4 = -0.87 + 0.0047 \cdot T_{cal} - 0.00036 \cdot n_L \cdot T_{cal} + 0.0024 \cdot h(n_L - 6) \cdot T_C \quad (5.5)$$

$$\hat{\gamma}_4 = -19 + 0.28 \cdot T_{cal} - 0.022 \cdot n_L \cdot T_{cal} + 0.16 \cdot h(n_L - 6) \cdot T_C \quad (5.6)$$

$$\hat{n}_{L4} = 6.8 - 0.014 \cdot T_{cal} + 0.0018 \cdot n_L \cdot T_{cal} + 0.0060 \cdot h(n_L - 6) \cdot T_C \quad (5.7)$$

$$\hat{v}_{cal,4} = 29 - 0.052 \cdot T_{cal} + 0.0011 \cdot n_L \cdot T_{cal} - 0.011 \cdot h(n_L - 6) \cdot T_C \quad (5.8)$$

The coefficients in equations (5.1) and (5.2) have the same signs and differ by a factor of roughly 100. This difference in magnitude fits the data well since the maxima of ρ and γ are separated by rough factor of 100 (see figure 5.5 and appendix B). The coefficients of the $v_{cal} \cdot T_{cal}$ interaction in equations (5.1) and (5.2) seem low, but the interaction has – considering the minimum value of the interaction of $300 \cdot 120 = 36\,000$ – several orders of magnitude higher influence on ρ_3 and γ_3 than the $h(6 - n_L)$ term. It is astonishing that the knot of the hinge function for equations (5.1)–(5.4) was chosen so low; basically only including the influence of the lowest number of coating

layers $n_L = 4$. Meaning for equations (5.1) and (5.2), that the lowest number of coating layer produces worse samples which is intuitive, but more than 6 layers do not improve the insulation. Encouragingly, the calcination heating rate v_{cal} (see equation (5.4)) has been predicted perfectly within numerical precision.

The coefficients of equations (5.5) and (5.6) show the same pattern as equations (5.1) and (5.2): identical signs and factor 100. The influence of T_{cal} on the measures of conductivity ($\hat{\rho}$ and $\hat{\gamma}$) is highly interesting. It was expected that calcination temperatures under 400 °C do not suffice to produce compact coating layers (corresponding to $+h(400 - T_{cal})$) and that therefore the resulting layer does not insulate well if T_{cal} is low (under 400 °C). The opposite is the case, though. *Optimizands* $\hat{\rho}$ and $\hat{\gamma}$ increase with higher calcination temperature according to equations (5.5) and (5.6) (see also figures 5.5c and 5.5d), ergo the resistance decreases and the conductivity increases with increasing calcination temperature contrary to expectations.

Furthermore, the coefficient of the T_{cal} term is the largest of all terms on $\hat{\rho}$ (5.5) and $\hat{\gamma}$ (5.6); interestingly also on \hat{n}_L (5.7) and \hat{v}_{cal} (5.8). The coefficient of the $n_L \cdot T_{cal}$ interaction is about a tenth in size of the T_{cal} coefficient, but has the extra factor n_L (range: 4–12), resulting in the products of coefficients and process variables being in the same order of magnitude for T_{cal} and $T_{cal} \cdot v_{cal}$ terms. It can be noted that the coefficient of the $n_L \cdot T_{cal}$ interaction always has contrary sign to T_{cal} coefficient in equations (5.5)–(5.8). This negative correlation could hint a compensation of the overestimated influence of T_{cal} on the *optimizands*. Similarly, signs of coefficients of $n_L \cdot T_{cal}$ and $h(n_L - 6) \cdot T_C$ correlate negatively in equations (5.5)–(5.8) and the product of coefficients and input variables are again similar in order of magnitude. The process variables T_{cal} and n_L appear as basis function (BF)s in generation 3 and 4, whereas v_{cal} only appears in generation 3's RFs and T_C only in generation 4's RFs. This can hint the importance of these variables in the process.

For each generation the MSE was calculated for which only samples from the optimization were used which were available at the time of prediction. That means 15, 20, 25 and 30 samples were used to calculate MSEs at $t = 1, 2, 3, 4$, respectively (compare with table 5.1). The MSEs are 64, 158, 54 and 50 for $t = 1, 2, 3, 4$, respectively. It is interesting that although prediction functions for $t = 3$ predicted v_{cal} perfectly, the combined MSE for $t = 4$ is lower. The high MSE at $t = 2$ can be explained by the prediction functions being only constant values. Apart from the second generation the MSE decreases over time, which indicates that the algorithm works. This decrease in MSE might be attributed to overfitting, though, since prediction and validation were performed on the same data. The MSE for each generation's prediction function was also calculated for pre-optimization (out-of-sample) samples, which are unseen data. The error sank again with each generation (except for the second generation): 102, 118, 58, 50, respectively for $t = 1, 2, 3, 4$. The decrease of MSE with out-of-sample data shows that the regularization method of the MARS algorithm in principle works on investigated samples and that decrease of MSE is not due to overfitting. When comparing the out-of-sample and in-sample MSEs it can be noted that at $t = 1$ the out-of-sample MSE is close to 1.5 fold of in-sample and at $t = 2$ vice versa. This shows the poor prediction ability at the beginning of the optimization which improved with generations. This decrease in validation (out-of-sample) MSE supports the hypothesis stated at the beginning of this section: the measure of conductivity was underestimated and the estimate improved over time (see table 5.2).

Even though the main *optimizands* were minimized as required, the RFs were not satisfactory. The two main problems were too many independent variables and too many dependent variables. Both seem closely related and overcomplicate the optimization, but both come with their own implications. Too many independent variables make it harder to distinguish variance due to random error (e.g. unmeasured and uncontrolled variables) from variance due to dependency. The difficulty of identifying meaningful correlation is mainly owed to the curse of dimensionality [84] which makes it hard to collect enough data for each dimension. The events per predictor variable (EPV) (data points per independent variable) are as low as 5 in the here presented optimization. By eliminating three independent variables the EPV could rise to 10, which is stated as rule of thumb for multivariate regression [109]. The results obtained with limited sample number deliver remarkable insight comparing to an EPV of 20–50 stated in the original MARS paper [82] and an EPV of around 20 in the original EMMA paper [80].

The main problem about too many dependent variables specific to this optimization procedure is that the same set of basis functions will be used to predict **all** *optimizands*. This leads to

competition between the BFs, as not all dependent variables may depend on the same independent variables. This effect is reinforced by the choice of two independent variables as dependent variables. Independent variables as dependent variables will likely be chosen in BFs, for including them in the RFs is an easy way to reduce the MSE and improve prediction accuracy. These BFs then "take away places" of BFs predicting other dependent variables in the RF. This in turn can lead to wrong predictions and therefore inefficient choice of future samples.

The two independent variables v_{cal} and n_L were included as *optimizands* for two reasons: Maximizing the calcination speed and minimizing the number of layer application iterations leads to minimization of process time. The process time optimization would be better placed in a follow up study. The second reason for including independent variables as *optimizands* was to check if the model works. This idea has two flaws: there is only one set of BFs for all *optimizands* and thus additional dependent variables not only complicate the model but make it more difficult to predict the actual dependent variables and chose future data points. Moreover, v_{cal} and n_L were only considered with 5% weight in the overall objective function. The second flaw is that there are obviously better methods to assess the quality of a model than to make it more complex. Validation methods for sparse data include re-sampling via leave-one-out or k -fold cross-validation [110].

5.3 Post-EMMA

In this section the experimental results are analyzed with different methods and compared with the optimization method EMMA which was integrated into the experiment selection process.

5.3.1 Analysis of Variance

The RFs produced by the last iteration of EMMA (see equations 5.5 and 5.6) put a high significance on the calcination temperature T_{cal} with regard to the measures of the conductance. ANOVA is used to double check this.

The null hypothesis is that the results were obtained by pure chance without any dependence of the insulation of a sample on the process variables. A one-way ANOVA was performed for every independent variable on ρ and γ . We used an α -value of 0.05 which means that the null hypothesis is rejected if the p-value is under 0.05. Usually, the α -value is chosen to be 0.01 or 0.05 [111, 112]. The choice of a relatively high *alpha*-value is motivated by the rather small amount of data and low events per variable ratio.

For both ρ and n_L (for both the EMMA dataset and the whole dataset (=EMMA+pre-EMMA)) the F-values for T_{cal} were in all cases under 0.01, indicating a real influence of T_{cal} on the conductance. Furthermore, both datasets exhibit a p-value < 0.05 for the interaction $T_{cal}:v_{cal}$ on ρ and the whole data set has p-value smaller than 0.01 of v_{cal} on ρ . Whereas v_{cal} on ρ for EMMA dataset has p-value of slightly over 0.05 ANOVA confirms that T_{cal} has influence on both γ and ρ and therefore agrees with EMMA. It is important to mention, that ANOVA assumes categorical data for inputs. Categorical inputs have no order and thus information is lost when using numerical data.

5.3.2 Linear Regression

The data obtained from the EMMA optimization was used to fit a linear model. The biggest difference between the linear model and EMMA is that the linear regression function will include all input variables, that the BFs are strictly linear and that there are no interaction terms. The coefficients of the n_L term have the same sign in equations 5.1, 5.2, 5.5, 5.6 (EMMA RFs) and 5.9, 5.10 and have comparable magnitudes too.

$$\begin{aligned} \hat{\rho} = & -0.022 \cdot c_{zr} - 0.022 \cdot n_L - 0.0029 \cdot v_C + 0.0058 \cdot T_C \\ & - 8.5 \cdot 10^{-5} \cdot v_{cal} + 0.0026 \cdot T_{cal} - 0.76 \end{aligned} \quad (5.9)$$

$$\begin{aligned} \hat{\gamma} = & 3.2 \cdot c_{zr} - 0.97 \cdot n_L - 1.2 \cdot v_C + 0.2 \cdot T_C \\ & - 0.014 \cdot v_{cal} + 0.13 \cdot T_{cal} + 7.2 \end{aligned} \quad (5.10)$$

All terms in equations 5.9 and 5.10 have the same sign except c_{zr} and the intercept. More particularly, T_{cal} and T_C have positive coefficients, i.e. a negative influence on the resistance and n_L ,

v_C and v_{cal} have negative coefficients, i.e. a positive influence on resistance. An indication that the data is strongly tainted by noise (see also section 5.3.1) is the low R^2 score of 0.41 of $\hat{\rho}$, which is (sad but true) even undercut by the R^2 of linear fit of $\hat{\gamma}$, 0.34. When the linear regression is trained on the whole data set, all terms have the same sign compared to the EMMA data set trained linear model. v_{cal} and n_L are perfectly predicted by linear regression as there is an exact direct proportionality and thus are not displayed here.

	c_{zr}	n_L	v_C	T_C	v_{cal}	T_{cal}
$\hat{\rho}$	2.04	46.58	1.15	9.21	13.50	27.52
$\hat{\gamma}$	10.24	7.10	16.46	10.97	7.68	47.55

Table 5.3: Influences of input variables on output variables in percent according to linear regression.

Regarding the coefficients of the linear model, the zirconium concentration in the starting solution c_{zr} and the number of layers applied n_L are largest in value, but the values which are factored with these coefficients - i.e. c_{zr} and n_L themselves – are the smallest among the input variables. Thus, even though T_{cal} might have a smaller coefficient in the linear RF, the influence on the dependent variable might be larger. For better comparison, the influence of an input variable was calculated by the formula: $I(x_i) = \frac{\bar{x}_i \cdot k_i}{\sum_i \bar{x}_i \cdot k_i}$ with $i = 1, \dots, d$ and where $\bar{x}_i = \frac{(x_{i,\min} + x_{i,\max})}{2}$ is average possible value of the i -th input variable and k_i is the coefficient of the i -th input variable in the linear regression. The biggest influence on $\hat{\rho}$ is n_L , followed by T_{cal} . While T_{cal} is the second largest influence on $\hat{\rho}$ as well, it is noteworthy that n_L contributes the least to $\hat{\gamma}$. That means that according to the linear regression, the number of coating layers increases the overall resistance of the zirconium ceramics while a lower calcination temperature leads to less defects. Interestingly, the number of coating layers has the least amount of influence on the number probability of defects, which seems to be an artifact of the noise rather than a process property. The two very different distributions of influence among the input variables can have three reasons: (1) the linear function is not a suitable regression function for the data distribution. (2) the signal to noise ratio is so low that the noise rather than the signal was fitted. (3) the *optimizands* indeed represent different properties of the material.

5.3.3 Grid Search, Kernel Ridge Regression and Support Vector Machine

The optimal parameters for each combination of *optimizand* (γ and ρ) and method (KRR and SVM) can be seen in table 5.4. The parameters are often extrema. The hyperparameter γ_{ml} took the smallest possible value (within the grid search) twice. In the polynomial kernel this leads to a lower influence of the input variables with respect to the intercept. The range of the hyperparameter α (see section 2.3.2) was 0.01–10 in the grid search. So, 2 and 10 are high values for the regularization parameter α of KRR. The hyperparameter ε specifies the size of the tube within which data does not contribute to the loss and the value 1 is the maximum for ε during this grid search. All these extreme values of hyperparameters stimulate regularization and indicate noisy and difficult to fit data. The R^2 values were calculated for each of the models using cross-validation. Resulting R^2 are all below zero (see table 5.4). A negative R^2 value means that the prediction provided by the model is worse than predicting only the mean of γ and ρ (see 4.1).

Finally, the MAE and MSE were calculated for different data sets (see table 5.5). Each models was trained on the EMMA data set. The first two columns ((e)) show statistics validated with data

method	R^2	kernel	degree	γ_{ml}	α	C	ε
KRR(γ)	-0.50 ± 0.25	poly	1	0.01	2	-	-
KRR(ρ)	-0.45 ± 0.52	rbf	-	0.2	10	-	-
SVM(γ)	-0.37 ± 0.39	poly	2	0.5	-	0.7	1.0
SVM(ρ)	-0.95 ± 0.44	poly	3	0.01	-	0.1	0.1

Table 5.4: R^2 values and best hyperparameters for each *optimizand*-method-combination

the model was trained with (30 data points) and thus, these columns contain the lowest numbers. The next two columns (p) were calculated with data obtained before the EMMA optimization (21 data points). As the pre-optimization data set was mainly used to find appropriate boundaries for the input variables, many data points are outside of the range of the EMMA data set. Thus, predictions for this data set are mainly extrapolations. It must be explicitly noted at this point that pre-EMMA samples differ from EMMA samples in the stabilization agent and therefore might deviate even more. The next two columns (c) were produced by using pre-EMMA data for validation, but only keeping points which sit inside the space spanned by the EMMA constraints (5 data points). It should be expected that these statistics should be lower than in the previous two columns. They are only lower for $MARS(\rho)$ and $LR(\rho)$, though. This is likely due to the small sample size. The last two columns show data for predictions of all available data (51 data points). These statistics are lower than the middle columns because of the inclusion of the training set in the test set.

In the self test of the first two columns MARS is the one method to rule them all and provides the lowest metrics. But as soon as unseen data is examined, linear regression excels by its simplicity.

γ	MAE(e)	MSE(e)	MAE(p)	MSE(p)	MAE(c)	MSE(c)	MAE(a)	MSE(a)
MARS	10	171	28	1084	30	1280	17	548
LR	13	250	24	747	34	1327	17	454
KRR	17	548	30	1477	35	2167	23	931
SVM	15	415	26	1050	28	1588	19	677
ρ	MAE(e)	MSE(e)	MAE(p)	MSE(p)	MAE(c)	MSE(c)	MAE(a)	MSE(a)
MARS	0.14	0.03	0.41	0.21	0.39	0.18	0.25	0.11
LR	0.19	0.06	0.30	0.15	0.38	0.13	0.24	0.09
KRR	0.25	0.12	0.38	0.24	0.43	0.29	0.31	0.17
SVM	0.23	0.13	0.41	0.28	0.45	0.33	0.31	0.20

Table 5.5: Comparison of MAE and MSE of different prediction methods for different data sets: EMMA data set (e), pre-EMMA data set (p), pre-EMMA data set within EMMA bounds (c) and complete dataset (a)

5.4 Hindsight Is Easier than Foresight

This section will discuss mistakes, improvements and lessons learned from each phase of this project. Firstly, the "Research, Exploration and Design Phase" will be discussed in subsection 5.4.1, followed by the "Narrowing, Pre-Optimization and Constraint Phase" reviewed in subsection 5.4.2 and finally the "Optimization, Modeling and Data Generation Phase" examined in subsection 5.4.3.

5.4.1 Research, Exploration and Design Phase

Scrutinize everything! At the beginning of a project it is important to get a "feel". I took the first recipe as only true recipe and tried to tune it to perfection. The problem was that the base line of that recipe was so low such that improving the recipe still lead to poor performance. This was only noticed after trying out another recipe. The author's tip here is to read as much and mostly diverse literature as possible and then try to mimic what ever can be. This will make it possible to acquire "feel" of what works and what does not. This is inspired by a quote from Thomas Torsney-Weir's data visualization classes: "create as many and different designs as possible". One should check what works, collect (dis-)advantages of methods and combine the best approaches. This can save you a lot of time.

Moreover, it is important to define the objective of the project as clear and unambiguous as possible. Again, try to reformulate the objective in multiple ways and choose the most fitting. Answer which question should be answered, and establish hypotheses with corresponding null

hypotheses. A general experimental design procedure was stated by Cherkassky and Mulier in “Learning from Data” [79]:

1. State the problem
2. Formulate the hypothesis
3. Design the experiment/data generation process
4. Collect the data and perform preprocessing
5. Estimate the model
6. Interpret the model/draw the conclusion

5.4.2 Narrowing, Pre-Optimization and Constraint Phase

The selection of input and output variables, data encoding/representation and incorporating a prior knowledge into the design of the learning system is often more critical for an overall success than the design of the learning machine itself [79]. After having decided on a base recipe, it is important to define limits and decrease the number of independent variables. The later is more important due to the curse of dimensionality [79], which states that with increasing dimensionality it get exponentially more difficult to arrive at the same density of data points. By excluding input variables from the model, the complexity of the model can be drastically reduced. This leads to more information, which can be extracted from the same amount of data, be it in more detail of the dependence or higher confidence interval. A good measure to compare data density with literature and other methods is the element per variable ratio. The EPV in this work is round 5, which is low in comparison with literature consulted [113–120]. Reducing variables is of course no easy task. How does one decide which variables to keep and which to incorporate into the model and which to discard? There is an intuitive (expert) way and a mathematical way. An expert would decide which variables to discard by his expert knowledge about the domain. Mathematical ways include principal component analysis and step-wise regression. Additionally, having wide limits can be an advantage when the data is noisy and the dependence is not too complicated, because by moving samples to the boundaries of the design space, the effect of the noise is reduced [121]. This suggest variable reduction favorable over limit narrowing. There are two opposing principles which should be kept in mind when reducing variables, though: By introducing a lot of variables into a model presupposes that they and their interactions (!) have a statistically significant effect [87]. On the other hand a model can be used to tell if a variable exerts an influence [37]. Another reason to avoid too many variables in a model is that they introduce more parameters to fit and overfitting happens in the regime where too many parameters are fitted with not enough data. Another step which is well placed in this phase is the estimate the measuring error by reproducing a single sample a certain number of times.

5.4.3 Optimization, Modeling and Data Generation Phase

During the optimization the main problems were too many in- and output variables, reproducibility and regularization. Variable selection has already been discussed in the previous section. When working with stochastic algorithms it is especially important to introduce reproducibility via a seed, which determines all subsequent pseudo-random numbers. Finally, during the EMMA optimization the input data were not regularized (scaled), although this could have increased the efficiency of the model. The input data could have been regularized to only take values $[-1; 1]$ by defining the regularized variable as $x' = \frac{x - \bar{x}}{x_{\max} - \bar{x}}$ with $\bar{x} = x_{\max} - x_{\min}$.

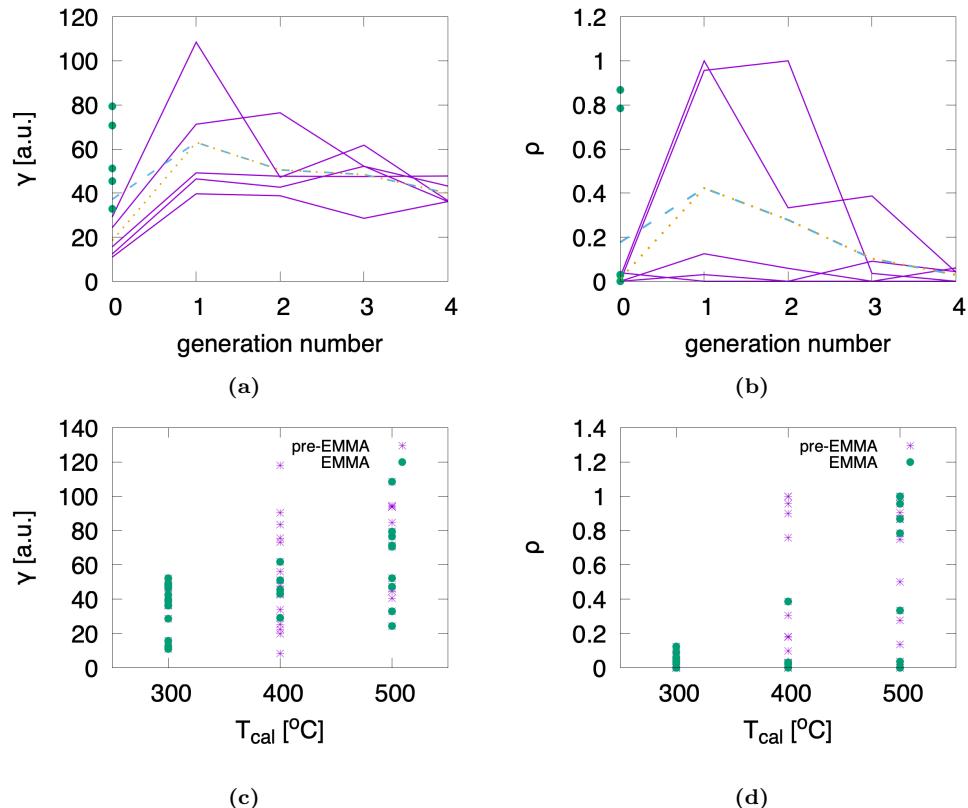


Figure 5.5: (a) calculated conductivity measures γ of each particle against the generation number (b) calculated pinhole ratio ρ of each particle against the generation number (c) calculated conductivity measures γ against calcination temperature T_{cal} (d) calculated pinhole ratio ρ against calcination temperature T_{cal} change colors

Chapter 6

Summary and Outlook

The goal of this work was to find a sol-gel blade coating process resulting in an electrical insulating zirconium ceramic on top of steel. This structure then can be used as substrate for thin layer photovoltaics. The ceramic needs to be homogeneous and defect free in order to insulate properly. A suitable recipe was found [96] and adapted. The recipe was improved by increasing the stability of the solution. The process variables were optimized via EMMA. The best parameters found by the EMMA optimization were $c_{zr} = 2$, $n_L = 10$, $v_C = 10$, $T_C = 40$, $v_{cal} = 120$ and $T_{cal} = 300$. After optimization, the data was analyzed with further machine learning methods like kernel ridge regression, support vector regression and linear regression. The comparison of methods showed that EMMA performed well on identifying the important correlation between conductance and process variables. A correlation of calcination temperature and conductance in the range of 300–500°C has been detected. Furthermore, the number of coating layers has been verified as an important factor. Next steps include to investigate if the correlation of conductance and calcination temperature are statistically significant or due to small sample size. The aging process could be examined via infrared spectroscopy.

Bibliography

- [1] F. J. Dyson, “Search for artificial stellar sources of infrared radiation,” Science, vol. 131, no. 3414, pp. 1667–1668, 1960. 1
- [2] N. S. Kardashev, “Transmission of information by extraterrestrial civilizations.,” Soviet Astronomy, Vol. 8, p. 217, vol. 8, p. 217, 1964. 1
- [3] P. S. Vasekar, A. H. Jahagirdar, and N. G. Dhere, “Photovoltaic characterization of copper–indium–gallium sulfide (cigs2) solar cells for lower absorber thicknesses,” Thin Solid Films, vol. 518, no. 7, pp. 1788–1790, 2010. Photovoltaics, solar energy materials and thin films - IMRC 2008, Cancun, Mexico. 1
- [4] J.-G. Yeo, Y.-G. Jung, and S.-C. Choi, “Design and microstructure of zro₂/sus316 functionally graded materials by tape casting,” Materials Letters, vol. 37, no. 6, pp. 304–311, 1998. 1
- [5] M. Michálek, G. Blugan, T. Graule, and J. Kuebler, “Comparison of aqueous and non-aqueous tape casting of fully stabilized zro₂ suspensions,” Powder Technology, vol. 274, pp. 276–283, 2015. 1
- [6] B. Petroleum, “Bp statistical review of world energy 2017,” Statistical review of world energy, vol. 65, 2017. 2.1
- [7] B. für Geowissenschaften und Rohstoffe, “Daten und entwicklungen der deutschen und globalen energieversorgung,” Statistical review of world energy, vol. 65, 2017. 2.1
- [8] D. Lübbert and F. Lange, “Uran als kernbrennstoff: Vorräte und reichweite,” Wissenschaftliche Dienste des deutschen Bundestages, Info-Brief WF VIII G–069/06, 2006. 2.1
- [9] M. Becquerel, “Mémoire sur les effets électriques produits sous l'influence des rayons solaires,” Comptes rendus hebdomadaires des séances de l'Académie des sciences, vol. 9, pp. 561–567, 1839. 2.1.1
- [10] W. Smith, “Effect of light on selenium during the passage of an electric current*,” Nature, vol. 7, p. 303, 1873. 2.1.1
- [11] W. G. Adams and R. E. Day, “V. the action of light on selenium,” Proceedings of the Royal Society of London, vol. 25, no. 171-178, pp. 113–117, 1877. 2.1.1
- [12] A. Einstein, “Über einem die erzeugung und verwandlung des lichtes betreffenden heuristischen gesichtspunkt,” Annalen der physik, vol. 4, 1905. 2.1.1
- [13] M. Prince and M. Wolf, “New developments in silicon photovoltaic devices,” Journal of the British Institution of Radio Engineers, vol. 18, no. 10, pp. 583–594, 1958. 2.1.1
- [14] R. L. Easton and M. J. Votaw, “Vanguard I IGY Satellite (1958 Beta),” Review of Scientific Instruments, vol. 30, pp. 70–75, 12 2004. 2.1.1
- [15] R. E. Stockwell, “Guidance for sputnik iii,” Ordnance, vol. 43, no. 232, pp. 673–675, 1959. 2.1.1

- [16] I. R. E. A. IRENA, Renewable power generation costs in 2022. eBook Partnership, 2022. 2.1.1
- [17] T. Markvart and L. Castañer, “Chapter ia-1 - principles of solar cell operation,” in Solar Cells (Second Edition) (A. McEvoy, L. Castañer, and T. Markvart, eds.), pp. 3–25, Elsevier, second edition ed., 2013. 2.1.2
- [18] O. Breitenstein, “Understanding the current-voltage characteristics of industrial crystalline silicon solar cells by considering inhomogeneous current distributions,” Opto-Electronics Review, vol. 21, no. 3, pp. 259–282, 2013. 2.1.2
- [19] H. Hahn, G. Frank, W. Klingler, A.-D. Meyer, and G. Störger, “Untersuchungen über ternäre chalcogenide. v. über einige ternäre chalcogenide mit chalkopyritstruktur,” Zeitschrift für anorganische und allgemeine Chemie, vol. 271, no. 3-4, pp. 153–170, 1953. 2.1.3
- [20] S. Wagner, J. Shay, P. Migliorato, and H. Kasper, “Cuinse2/cds heterojunction photovoltaic detectors,” Applied Physics Letters, vol. 25, no. 8, pp. 434–435, 1974. 2.1.3
- [21] L. Kazmerski, F. White, and G. Morgan, “Thin-film cuinse2/cds heterojunction solar cells,” Applied Physics Letters, vol. 29, no. 4, pp. 268–270, 1976. 2.1.3
- [22] M. A. Green, E. D. Dunlop, G. Siefer, M. Yoshita, N. Kopidakis, K. Bothe, and X. Hao, “Solar cell efficiency tables (version 61),” Progress in Photovoltaics: Research and Applications, vol. 31, no. 1, pp. 3–16, 2023. 2.1.3
- [23] K. Mertens, Photovoltaik: Lehrbuch zu Grundlagen, Technologie und Praxis. Carl Hanser Verlag GmbH Co KG, 2020. 2.1, 2.2
- [24] U. Rau and H. Schock, “Chapter ic-3 - cu(in,ga)se2 thin-film solar cells,” in Solar Cells (Second Edition) (A. McEvoy, L. Castañer, and T. Markvart, eds.), pp. 261–304, Elsevier, second edition ed., 2013. 2.1
- [25] J. Hedstrom, H. Ohlsen, M. Bodegard, A. Kyller, L. Stolt, D. Hariskos, M. Ruckh, and H.-W. Schock, “Zno/cds/cu(in,ga)se/sub 2/ thin film solar cells with improved performance,” in Conference Record of the Twenty Third IEEE Photovoltaic Specialists Conference - 1993 (Cat. No.93CH3283-9), pp. 364–371, 1993. 2.1.3
- [26] T. Feurer, P. Reinhard, E. Avancini, B. Bissig, J. Löckinger, P. Fuchs, R. Carron, T. P. Weiss, J. Perrenoud, S. Stutterheim, S. Buecheler, and A. N. Tiwari, “Progress in thin film cigs photovoltaics – research and development, manufacturing, and applications,” Progress in Photovoltaics: Research and Applications, vol. 25, no. 7, pp. 645–667, 2017. 2.1.3
- [27] M. A. Anwar, T. Kurniawan, Y. P. Asmara, W. S. W. Harun, A. N. Oumar, and A. B. D. Nandyanto, “Morphology evaluation of ZrO₂dip coating on mild steel and its corrosion performance in NaOH solution,” IOP Conference Series: Materials Science and Engineering, vol. 257, p. 012087, oct 2017. 2.2.1, 3.2, 5.1
- [28] K. Kukli, K. Forsgren, M. Ritala, M. Leskelä, J. Aarik, and A. Haärsta, “Dielectric properties of zirconium oxide grown by atomic layer deposition from iodide precursor,” Journal of The Electrochemical Society, vol. 148, no. 12, p. F227, 2001. 2.2.1
- [29] R. H. Nielsen and G. Wilfing, Zirconium and Zirconium Compounds. John Wiley & Sons, Ltd, 2010. 2.2.1
- [30] H. Yang and P. Jiang, “Large-scale colloidal self-assembly by doctor blade coating,” Langmuir, vol. 26, no. 16, pp. 13173–13182, 2010. 2.2.2
- [31] S. Swann, “Magnetron sputtering,” Physics in Technology, vol. 19, pp. 67–75, mar 1988. 2.2.3
- [32] M. Kaliva and M. Vamvakaki, “Chapter 17 - nanomaterials characterization,” in Polymer Science and Nanotechnology (R. Narain, ed.), pp. 401–433, Elsevier, 2020. 2.2.4, 2.2.5

- [33] K. Vernon-Parry, “Scanning electron microscopy: an introduction,” III-Vs Review, vol. 13, no. 4, pp. 40–44, 2000. 2.2.4
- [34] G. Schwedt, Analytische chemie: Grundlagen, Methoden und Praxis. John Wiley & Sons, 2008. 2.2.4
- [35] B. C. Healy, “Machine and deep learning in ms research are just powerful statistics – no,” Multiple Sclerosis Journal, vol. 27, no. 5, pp. 663–664, 2021. PMID: 33625302. 2.3
- [36] A. Gontcharov, “No, ai is not statistics,” 2019. Last Accessed 2022-06-21. 2.3
- [37] G. Härtler, Statistisch gesichert und trotzdem falsch? Springer Spektrum, 2014. 2.3, 5.4.2
- [38] J. Howard, “Artificial intelligence: Implications for the future of work,” American Journal of Industrial Medicine, vol. 62, no. 11, pp. 917–926, 2019. 2.3.1
- [39] M. J, M. ML, R. N, and S. CE, “A proposal for the dartmouth summer research project on artificial intelligence,” 1955. 2.3.1
- [40] P. McCorduck and C. Cfe, Machines who think: A personal inquiry into the history and prospects of artificial intelligence. CRC Press, 1982. 2.3.1
- [41] M. A. Apter, “Machines who think: A personal inquiry into the history and prospects of artificial intelligence by pamela mccorduck,” Leonardo, vol. 15, no. 3, p. 242, 1982. 2.3.1
- [42] D. Silver, T. Hubert, J. Schrittwieser, I. Antonoglou, M. Lai, A. Guez, M. Lanctot, L. Sifre, D. Kumaran, T. Graepel, T. Lillicrap, K. Simonyan, and D. Hassabis, “Mastering chess and shogi by self-play with a general reinforcement learning algorithm,” 2017. 2.3.1
- [43] F.-H. Hsu, “Ibm’s deep blue chess grandmaster chips,” IEEE Micro, vol. 19, no. 2, pp. 70–81, 1999. 2.3.1
- [44] A. M. Turing, “The imitation game,” Mind, vol. 59, no. 236, pp. 433–460, 1950. 2.3.1
- [45] GPT-3, “A robot wrote this entire article. are you scared yet, human?.” Last Accessed 2022-06-14. 2.3.1
- [46] E. A. M. van Dis, J. Bollen, W. Zuidema, R. van Rooij, and L. Bockting, “Chatgpt: five priorities for research,” 2023. 2.3.1
- [47] J. R. Searle, “Minds, brains, and programs,” Behavioral and Brain Sciences, vol. 3, no. 3, p. 417–424, 1980. 2.3.1
- [48] J. Searle, “I married a computer,” The New York Review of Books, vol. 8, 1999. 2.3.1
- [49] N. Tiku, “The google engineer who thinks the company’s ai has come to life,” The Washington Post, vol. 11, 2022. 2.3.1
- [50] M. Campbell, A. Hoane, and F. hsiung Hsu, “Deep blue,” Artificial Intelligence, vol. 134, no. 1, pp. 57–83, 2002. 2.3.1
- [51] L.-F. Li, X. Wang, W.-J. Hu, N. N. Xiong, Y.-X. Du, and B.-S. Li, “Deep learning in skin disease image recognition: A review,” IEEE Access, vol. 8, pp. 208264–208280, 2020. 2.3.1
- [52] R. C. Deo, “Machine learning in medicine,” Circulation, vol. 132, no. 20, pp. 1920–1930, 2015. 2.3.1
- [53] E. J. Topol, “High-performance medicine: the convergence of human and artificial intelligence,” Nature medicine, vol. 25, no. 1, pp. 44–56, 2019. 2.3.1
- [54] H. Fujiyoshi, T. Hirakawa, and T. Yamashita, “Deep learning-based image recognition for autonomous driving,” IATSS Research, vol. 43, no. 4, pp. 244–252, 2019. 2.3.1

- [55] J. Westermayr, M. Gastegger, M. F. S. J. Menger, S. Mai, L. González, and P. Marquetand, “Machine learning enables long time scale molecular photodynamics simulations,” *Chem. Sci.*, vol. 10, pp. 8100–8107, 2019. 2.3.1
- [56] G. B. Goh, C. Siegel, A. Vishnu, N. O. Hodas, and N. Baker, “Chemception: A deep neural network with minimal chemistry knowledge matches the performance of expert-developed qsar/qspr models,” 2017. 2.3.1
- [57] D. Jha, L. Ward, A. Paul, W.-k. Liao, A. Choudhary, C. Wolverton, and A. Agrawal, “Elemnet: Deep learning the chemistry of materials from only elemental composition,” *Scientific reports*, vol. 8, no. 1, pp. 1–13, 2018. 2.3.1
- [58] I. H. Sarker, M. H. Furhad, and R. Nowrozy, “Ai-driven cybersecurity: An overview, security intelligence modeling and research directions,” *SN Computer Science*, vol. 2, p. 173, 2021. 2.3.1
- [59] T. M. Andrews, 2017. Last Accessed on 2022-06-03. 2.3.1
- [60] M. L. Littman, I. Ajunwa, G. Berger, C. Boutilier, M. Currie, F. Doshi-Velez, G. Hadfield, M. C. Horowitz, C. Isbell, H. Kitano, K. Levy, T. Lyons, M. Mitchell, J. Shah, S. Sloman, S. Vallor, and T. Walsh, “Gathering strength, gathering storms: The one hundred year study on artificial intelligence (ai100) 2021 study panel report.,” 2021. 2.3.1
- [61] M. V. Koroteev, “Bert: A review of applications in natural language processing and understanding,” 2021. 2.3.1
- [62] X. Liu, Y. Zheng, Z. Du, M. Ding, Y. Qian, Z. Yang, and J. Tang, “Gpt understands, too,” 2021. 2.3.1, 1
- [63] J. Parviainen and M. Coeckelbergh, “The political choreography of the sophia robot: beyond robot rights and citizenship to political performances for the social robotics market,” *AI & SOCIETY*, vol. 36, pp. 715–724, 2021. 2.3.1
- [64] H. Mansourifar and W. Shi, “One-shot gan generated fake face detection,” 2020. 2.3.1
- [65] G. Marcus, E. Davis, and S. Aaronson, “A very preliminary analysis of dall-e 2,” 2022. 2.3.1
- [66] M. Guzdial and M. Riedl, “Toward game level generation from gameplay videos,” 2016. 2.3.1
- [67] last accessed 2022-06-13. 2.3.1
- [68] O. Hassan, “Artificial intelligence, neom and saudi arabia’s economic diversification from oil and gas,” *The Political Quarterly*, vol. 91, no. 1, pp. 222–227, 2020. 2.3.1
- [69] H. Chiroma, U. A. Abdullahi, S. M. Abdulhamid, A. Abdulsalam Alarood, L. A. Gabralla, N. Rana, L. Shuib, I. A. Targio Hashem, D. E. Gbenga, A. I. Abubakar, A. M. Zeki, and T. Herawan, “Progress on artificial neural networks for big data analytics: A survey,” *IEEE Access*, vol. 7, pp. 70535–70551, 2019. 2.3.2
- [70] C. M. Bishop, “Neural networks and their applications,” *Review of scientific instruments*, vol. 65, no. 6, pp. 1803–1832, 1994. 2.3.2
- [71] Y. LeCun, Y. Bengio, et al., “Convolutional networks for images, speech, and time series,” *The handbook of brain theory and neural networks*, vol. 3361, no. 10, p. 1995, 1995. 2.3.2
- [72] J. Westermayr, M. Gastegger, and P. Marquetand, “Combining schnet and sharc: The schnarc machine learning approach for excited-state dynamics,” *The journal of physical chemistry letters*, vol. 11, no. 10, pp. 3828–3834, 2020. 2.3.2
- [73] M. Rupp, “Machine learning for quantum mechanics in a nutshell,” *International Journal of Quantum Chemistry*, vol. 115, no. 16, pp. 1058–1073, 2015. 2.3.2
- [74] C. Rudin, “Least squares and friends,” 2020. last accessed 2022-11-08. 2.3.2

- [75] B. E. Boser, I. M. Guyon, and V. N. Vapnik, “A training algorithm for optimal margin classifiers,” in *Proceedings of the Fifth Annual Workshop on Computational Learning Theory*, COLT ’92, (New York, NY, USA), p. 144–152, Association for Computing Machinery, 1992. 2.3.2
- [76] C. Cortes and V. Vapnik, “Support-vector networks,” *Machine learning*, vol. 20, pp. 273–297, 1995. 2.3.2, 2.3.2
- [77] H. Drucker, C. J. Burges, L. Kaufman, A. Smola, and V. Vapnik, “Support vector regression machines,” *Advances in neural information processing systems*, vol. 9, 1996. 2.3.2
- [78] P. H. Winston, *Artificial Intelligence*. Addison-Wesley Longman Publishing Co., Inc., 1992. 2.3.2
- [79] V. Cherkassky and F. Mulier, *Learning from Data*. Wiley, 1998. 2.3.2, 2.3.5, 5.4.1, 5.4.2
- [80] L. Villanova, P. Falcaro, D. Carta, I. Poli, R. Hyndman, and K. Smith-Miles, “Functionalization of microarray devices: Process optimization using a multiobjective pso and multiresponse mars modeling,” in *IEEE Congress on Evolutionary Computation*, pp. 1–8, IEEE, 2010. 2.3.3, 4.1, 5.2
- [81] J. Kennedy and R. Eberhart, “Particle swarm optimization,” in *Proceedings of ICNN’95-international conference on neural networks*, vol. 4, pp. 1942–1948, IEEE, 1995. 2.3.3, 4.1
- [82] J. H. Friedman, “Multivariate adaptive regression splines,” *Annals of Statistics*, vol. 19, pp. 1 – 67, March 1991. 2.3.4, 5.2
- [83] last accessed 2022-10-17. 2.3.4
- [84] J. H. Friedman, “Fitting functions to noisy data in high dimensions,” *Department of Statistics-Stanford University Technical Report*, vol. 101, pp. 1–36, 1988. 2.3.4, 2.3.4, 5.2
- [85] G. Wahba and P. Craven, “Smoothing noisy data with spline functions. estimating the correct degree of smoothing by the method of generalized cross-validation.,” *Numerische Mathematik*, vol. 31, pp. 377–404, 1979. 2.3.4
- [86] G. C. Goodwin and R. L. Payne, “Chapter 6 experiment design,” in *Dynamic System Identification* (G. C. Goodwin and R. L. Payne, eds.), vol. 136 of *Mathematics in Science and Engineering*, pp. 124–174, Elsevier, 1977. 2.3.5
- [87] R. F. Gunst and R. L. Mason, “Fractional factorial design,” *WIREs Computational Statistics*, vol. 1, no. 2, pp. 234–244, 2009. 2.3.5, 5.4.2
- [88] K. Vanaja and R. S. Rani, “Design of experiments: Concept and applications of plackett burman design,” *Clinical Research and Regulatory Affairs*, vol. 24, no. 1, pp. 1–23, 2007. 2.3.5
- [89] A. Miller and R. R. Sitter, “Using the folded-over 12-run plackett—burman design to consider interactions,” *Technometrics*, vol. 43, no. 1, pp. 44–55, 2001. 2.3.5
- [90] J. C. Wang and C. F. J. Wu, “A hidden projection property of plackett-burman and related designs,” *Statistica Sinica*, vol. 5, no. 1, pp. 235–250, 1995. 2.3.5
- [91] F. A. C. Viana, “A tutorial on latin hypercube design of experiments,” *Quality and Reliability Engineering International*, vol. 32, no. 5, pp. 1975–1985, 2016. 2.3.5
- [92] U. M. Diwekar and J. R. Kalagnanam, “Efficient sampling technique for optimization under uncertainty,” *AICHE Journal*, vol. 43, no. 2, pp. 440–447, 1997. 2.3.5
- [93] R. A. Fisher et al., “On the ”probable error” of a coefficient of correlation deduced from a small sample.,” *Metron*, vol. 1, pp. 3–32, 1921. 2.3.6

- [94] T. YAMASHITA, R. YAMAUCHI, and K. ANDO, “Progress in ultrasonic cleaning research,” *JAPANESE JOURNAL OF MULTIPHASE FLOW*, vol. 32, no. 2, pp. 210–217, 2018. 3.1
- [95] S. Awad and R. Nagarajan, “Chapter 6 - ultrasonic cleaning,” in *Developments in Surface Contamination and Cleaning* (R. Kohli and K. L. Mittal, eds.), pp. 225–280, Oxford: William Andrew Publishing, 2010. 3.1
- [96] B. Hu, E. Jia, B. Du, and Y. Yin, “A new sol-gel route to prepare dense al₂o₃ thin films,” *Ceramics International*, vol. 42, no. 15, pp. 16867–16871, 2016. <https://www.sciencedirect.com/science/article/pii/S0272884216312548>. 3.2, 3.2.2, 3.2.2, 5.1, 6
- [97] R. Plackett and L. Burman, “The design of optimal multifactorial experiments,” *Biometrika*, vol. 33, pp. 305–325, 1946. 3.7
- [98] L. Breiman and J. H. Friedman, “Predicting multivariate responses in multiple linear regression,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 59, no. 1, pp. 3–54, 1997. 4.1
- [99] D. Carta, L. Villanova, S. Costacurta, A. Patelli, I. Poli, S. Vezzu, P. Scopece, F. Lisi, K. Smith-Miles, R. J. Hyndman, et al., “Method for optimizing coating properties based on an evolutionary algorithm approach,” *Analytical chemistry*, vol. 83, no. 16, pp. 6373–6380, 2011. 4.1
- [100] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, et al., “Scikit-learn: Machine learning in python,” *the Journal of machine Learning research*, vol. 12, pp. 2825–2830, 2011. 4.2
- [101] G. Katz, “X-ray diffraction powder pattern of metastable cubic zro₂,” *Am. Ceram. Soc.*, vol. 54, p. 531, 1971. 5.1.1, 5.2
- [102] R. Purohit, S. Saha, and A. Tyagi, “Combustion synthesis of nanocrystalline zro₂ powder: Xrd, raman spectroscopy and tem studies,” *Materials Science and Engineering: B*, vol. 130, no. 1-3, pp. 57–60, 2006. 5.1.1
- [103] D. J. Dumin, “Measurement of film thickness using infrared interference,” *Review of Scientific Instruments*, vol. 38, no. 8, pp. 1107–1109, 1967. 5.1.2
- [104] J. B. de Lima Filho and Ángel Alberto Hidalgo, “Film thickness by interference pattern and optical characterization of polyaniline by spectroscopic ellipsometry,” *Synthetic Metals*, vol. 223, pp. 80–86, 2017. 5.1.2
- [105] A. Pérez-Tomás, A. Lima, Q. Billon, I. Shirley, G. Catalan, and M. Lira-Cantú, “A solar transistor and photoferroelectric memory,” *Advanced Functional Materials*, vol. 28, no. 17, p. 1707099, 2018. 5.1.3
- [106] A. Sinhamahapatra, J.-P. Jeon, J. Kang, B. Han, and J.-S. Yu, “Oxygen-deficient zirconia (zro_{2-x}): A new material for solar light absorption,” *Scientific Reports*, vol. 6, 2016. 5.1.3
- [107] R. M. Feenstra, “Scanning tunneling spectroscopy,” *Surface science*, vol. 299, pp. 965–979, 1994. 5.1.3
- [108] S. Datta, W. Tian, S. Hong, R. Reifenberger, J. I. Henderson, and C. P. Kubiak, “Current-voltage characteristics of self-assembled monolayers by scanning tunneling microscopy,” *Physical Review Letters*, vol. 79, no. 13, p. 2530, 1997. 5.1.3
- [109] E. Vittinghoff and C. E. McCulloch, “Relaxing the rule of ten events per variable in logistic and cox regression,” *American journal of epidemiology*, vol. 165, no. 6, pp. 710–718, 2007. 5.2
- [110] N. Kartam, I. Flood, and J. H. Garrett, “Validation and verification,” in *Artificial neural networks for civil engineers: fundamentals and applications*, ASCE, 1997. 5.2

- [111] R. Hoffman, “The concept of statistical significance,” 2020. last accessed 2022-11-08. 5.3.1
- [112] T. Sellke, M. J. Bayarri, and J. O. Berger, “Calibration of p-values for testing precise null hypotheses,” *The American Statistician*, vol. 55, no. 1, pp. 62–71, 2001. 5.3.1
- [113] M. A. Ahmed Al-dujaili, S. Jaheel, and H. N. Abbas, “Preparation of ha/ β -tcp scaffold and mechanical strength optimization using a genetic algorithm method,” *Journal of the Australian Ceramic Society*, vol. 53, pp. 41–48, 2017. 5.4.2
- [114] V. R. Fernandes, C. M. Vicente, N. Wada, P. S. André, and R. A. Ferreira, “Multi-objective genetic algorithm applied to spectroscopic ellipsometry of organic-inorganic hybrid planar waveguides,” *Optics Express*, vol. 18, no. 16, pp. 16580–16586, 2010. 5.4.2
- [115] A. Kaczmarowski, S. Yang, I. Szlufarska, and D. Morgan, “Genetic algorithm optimization of defect clusters in crystalline materials,” *Computational Materials Science*, vol. 98, pp. 234–244, 2015. 5.4.2
- [116] M. Mahdavi Jafari, S. Soroushian, and G. R. Khayati, “Hardness optimization for al6061-mwcnt nanocomposite prepared by mechanical alloying using artificial neural networks and genetic algorithm,” *Journal of Ultrafine Grained and Nanostructured Materials*, vol. 50, no. 1, pp. 23–32, 2017. 5.4.2
- [117] S. Soltanali, R. Halladj, S. Tayyebi, and A. Rashidi, “Neural network and genetic algorithm for modeling and optimization of effective parameters on synthesized zsm-5 particle size,” *Materials Letters*, vol. 136, pp. 138–140, 2014. 5.4.2
- [118] R. Panwar, V. Agarwala, and D. Singh, “Design and experimental verification of a thin broadband nanocomposite multilayer microwave absorber using genetic algorithm based approach,” *AIP Conference Proceedings*, vol. 1620, no. 1, pp. 406–415, 2014. 5.4.2
- [119] M. F. Schubert, F. W. Mont, S. Chhajed, D. J. Poxson, J. K. Kim, and E. F. Schubert, “Design of multilayer antireflection coatings made from co-sputtered and low-refractive-index materials by genetic algorithm,” *Optics express*, vol. 16, no. 8, pp. 5290–5298, 2008. 5.4.2
- [120] A. Shanaghi, A. Sabour Rouhaghdam, S. A. Tabatabaei, M. Shariatpanahi, A. Rastgoo, A. K. Yousefi, and M. Samadzadeh, “An experimental investigation of corrosion resistance of mild steel by sol-gel process using tio₂ nanostructure coating and prediction of optimal parameters,” *Proceedings of the Institution of Mechanical Engineers, Part N: Journal of Nanoengineering and Nanosystems*, vol. 227, no. 2, pp. 105–111, 2013. 5.4.2
- [121] A. Giunta, S. Wojtkiewicz, and M. Eldred, “Overview of modern design of experiments methods for computational simulations (invited),” *41st Aerospace Sciences Meeting and Exhibit*, 2003. 5.4.2

Appendix A

Pre-EMMA Experiments

gn	nr	enr	c_{zr}	λ	v_C	T_C	v_{cal}	T_{cal}	γ	ρ
-	146	-	1	10	10	20	2	400	33.82319	0.09804
-	150	-	3	5	10	20	2	400	22.45401	0.02941
-	150	-	3	5	10	20	2	400	8.389750	0.00000
-	151	-	3	2	10	20	2	400	42.35738	0.18182
-	152	-	4	2	10	20	2	400	47.78915	0.18000
-	153	-	4	3	10	20	2	400	19.82546	0.00000
-	154	-	4	4	10	60	2	400	25.19964	0.00000
-	156	-	2	3	10	60	2	400	73.22186	0.75862
-	157	-	2	3	10	60	2	400	83.55425	1.00000
-	158	-	2	3	10	60	2	400	55.98166	0.30612
-	160	-	3	4	10	60	2	400	118.11504	1.00000
<hr/>										
-	186	-	2	5	10	20	2	400	75.53455	0.95652
-	187	-	1	10	5	50	2	400	43.48684	0.02222
-	188	-	1	4	0.1	20	2	500	40.46723	0.27586
-	190	-	5	10	0.1	20	2	500	93.74032	0.90476
-	192	-	4	6	14	60	4	500	44.52547	0.13636
-	194	-	5	4	5	20	8	400	90.27868	0.90000
-	195	-	1	5	5	80	2	500	84.67789	0.75000
-	198	-	5	10	5	80	8	500	94.74563	0.86486
-	199	-	4	6	14	60	4	500	93.97734	0.97368
-	201	-	4	6	14	60	4	500	46.08935	0.50000

Table A.1: gn – generation number; nr – experiment number; enr – EMMA serial number

Appendix B

EMMA Experiments

gn	nr	enr	c_{zr}	λ	v_C	T_C	v_{cal}	T_{cal}	γ	ρ
1	207	6113	2	10	20	40	2	500	24.32170	0.00000
1	208	2850	3	8	18	70	18	300	12.38955	0.00000
1	209	5526	3	6	10	50	18	400	29.16378	0.02083
1	211	6554	3	10	16	80	2	500	70.64908	0.78431
1	212	2947	4	6	16	80	18	300	10.95940	0.00000
1	213	8318	3	12	12	80	14	500	79.39680	0.86842
1	214	5648	5	6	10	60	18	400	51.09902	0.03030
1	217	7374	3	10	14	50	10	500	32.86985	0.00000
1	216	2700	5	12	14	60	18	300	15.64646	0.03846
1	218	3956	5	10	20	60	6	400	45.50493	0.02778
2	220	6101	2	4	20	40	2	500	71.25290	0.95652
2	222	7201	2	4	10	40	10	500	108.47050	1.00000
2	223	1462	3	4	12	60	10	300	39.68027	0.12500
2	224	2883	4	4	10	80	18	300	46.44359	0.03030
2	225	1680	5	12	20	70	10	300	49.18166	0.00000
3	226	0001	2	4	10	40	2	300	38.79620	0.05882
3	227	6001	2	4	10	40	2	500	47.17184	0.33333
3	228	6102	3	4	20	40	2	500	76.44246	1.00000
3	229	2884	5	4	10	80	18	300	42.66628	0.00000
3	231	0360	5	12	20	60	2	300	47.64474	0.00000
4	234	4202	3	4	10	40	10	400	61.73920	0.38710
4	235	6105	2	6	20	40	2	500	52.13835	0.03571
4	237	1500	5	12	14	60	10	300	47.53164	0.00000
4	238	1486	3	6	14	60	10	300	28.56022	0.00000
4	239	2923	4	4	14	80	18	300	52.18121	0.09091
5	240	2971	4	8	18	80	18	300	36.17699	0.06061
5	241	2530	3	8	10	50	18	300	36.54716	0.04167
5	242	3077	2	12	16	40	2	400	43.19611	0.00000
5	243	2733	2	10	18	60	18	300	36.16643	0.04348
5	244	2535	4	10	10	50	18	300	47.75640	0.00000

Table B.1: gn – generation number; nr – experiment number; enr – EMMA serial number