

Gerald Birngruber, B.Sc.

# Digital Energy Twin: A data-driven approach to analyze and optimize industrial energy systems

## Master's Thesis

to achieve the university degree of

Diplom-Ingenieur

Master's degree programme: Computer Science

submitted to

## Graz University of Technology

Institute for Softwaretechnology

Supervisor Univ.-Prof. Dipl.-Ing. Dr.techn. Wotawa Franz

Co-Supervisor Mag.phil. Dr.techn. MA MA Schweiger Gerald

Graz, April 2021

# Affidavit

I declare that I have authored this thesis independently, that I have not used other than the declared sources/resources, and that I have explicitly indicated all material which has been quoted either literally or by content from the sources used. The text document uploaded to TUGRAZonline is identical to the present master's thesis.

Date Signature

# Acknowledgements

I would like to thank my supervisor Franz Wotawa for the support by correcting my thesis and providing feedback for the work. A special thanks goes to Gerald Schweiger, my Co-Supervisor. He employed me in his research group and gave me tremendous background knowledge for the thesis. Further, he established connections to project partners of the Digital Energy Twin (DET) project. He spent endless hours correcting my thesis drafts, answering my questions, and supporting me wherever possible. Special mention goes to the members of my research group. Thomas Schranz and Thomas Schwengler helped me with my entrance into research and gave generous support in solving technical issues. A special thanks also goes to my colleague and lab partner Michael Grömer. With his help, it was possible to organize and perform the lab experiment at the TU Graz physics department. He also had good ideas for the data model creation. He was very pleasant to work with.

The lab experiment had some more supporters. In the first place, there was Markus Koch, who provided a laboratory room and equipment for the experimentation. He also helped with ideas for funding the experiment. The money was then offered by the NAWI Graz organization, which has a budget for student projects. Thank you, NAWI Graz, for the trust in our work and the funding. Some loaned equipment we borrowed from my former employers Greiner Bio-One Diagnostics GmbH and Genspeed Biotech GmbH. The equipment lent by you made our experiment possible in the first place, allowed us to do more precise measurements, and saved our budget.

The last words of gratitude go to the project partners of the DET project and first of all to the organizantion unit for the whole project AEE IN-TEC. Ahmed Junaid Tahir and Carles Ribas Tugores are enthusiastic about

pushing the project forward. Many thanks to Carles Ribas Tugores for supporting me in the modeling part of the thesis and always took the time to answer my questions and debug my created models. Many thanks also to Ahmed Junaid Tahir for the organization of all the meetings, workshops, and submissions. He always tried to support me, and together we found the right solutions for upcoming problems. Another valuable partner is ENERTEC Naftz & Partner GmbH & Co KG, especially Eugen Naftz and Angelika Swatek. They have excellent knowledge about the processes and machines in the electroplating plant and spent hours patiently explaining many details for the modeling part.

# <span id="page-4-0"></span>Abstract

Climate change affects everyone. A primary culprit is  $CO<sub>2</sub>$  emission. Reduction of  $CO<sub>2</sub>$  emission could curb climate change. Process optimization can save energy and thereby reduce  $CO<sub>2</sub>$ emission. In Industry 4.0, a digital twin (DT) is a key-enabler for process optimization. This work is part of an Austrian Research Promotion Agency (FFG) project. The FFG project aims to build a digital twin for an electroplating plant. Within this thesis, the main goals are the improvement, simplification, and extension of existing parts. This thesis consists of two parts. The first part is the extension of a DT model with control mechanisms for machines and processes. The second part describes the development of data models. The data models deliver data as input for the digital twin model. In the first place, the DT needs an approximation for the electroplating process. A lab experiment took place to collect data about the process. Data models are generated with the collected data, which delivers the energy consumption depending on various input parameters. Besides the electroplating, there are a lot of other components. The plant provides a massive amount of sensor data. However, it is too much data; thus, the computations for the DT get too time-consuming. An automated function-based approximation of the data solves the problem. The DT is implemented in Modelica. Modelica is a multidomain modeling language for physical models. The results of the modeling part are control extensions for the existing machines and processes. Various programming languages, such as Python, can provide input for the Modelica models. Thus all approximations and data models are Python programs. The lab experiment took place at the TU Graz physics department. The data models use various machine learning algorithms like polynomial regression, decision tree regression, random forest regression, and neural networks to generate input for the Modelica model. There is also a discussion about the machine learning methods and their performance to handle a problem and when it is better to find another prediction method, such as function-based

approximation. It uses optimizer functions to fit the given functions best to real data and return the functions' parameters. Thus the function predicts future events. The result is a framework that creates predictions for machine data in arbitrary time frames. Altogether the separately developed parts work well together. With the optimizations, the computations are faster. The predictions are not as exact as real data, but they approximate the real data well, and therefore, it is an excellent trade to lose a little precision but receive better computation times. This leads to faster development and has benefits for the progress of the whole project.

# **Contents**





# <span id="page-8-0"></span>1. Introduction

*Die Klimakrise ist im vergangenen Jahr durch die Pandemie aus den Schlagzeilen gerutscht. Sie ist aber deswegen nicht verschwunden - im Gegenteil. [. . . ]*

*Wir stehen vor der großten Herausforderung der Menschheit in diesem ¨ Jahrtausend. Einer großeren Herausforderung als jener durch die Pan- ¨ demie. Und gegen die Klimakrise wird es keine Impfung geben. Da sind wir alle selbst gefordert. Es wird ein Wettlauf mit der Zeit. Aber wir konnen ihn gewinnen. Wir m ¨ ussen ihn gewinnen, wollen wir diesen ¨ Planeten fur uns Menschen gut bewohnbar halten. ¨* - Alexander Van der Bellen (president of austria) in Austria's Virtual New Year's Reception 2021

Due to the presence of Sars-CoV-2 in the daily news, climate change has moved into the background again, but it is still essential and will affect our daily lives more and more - a few examples: Food prices rise because raw ingredients, like wheat, are threatened by heatwaves (Hasegawa et al., [2018](#page-87-0)). The number of natural disasters rose in the past few years (Wuebbles et al., [2017](#page-89-0)). Drinking water is becoming less and less, or more challenging to tap (Watts et al., [2016](#page-89-1)).

Greenhouse gases (GHG) are responsible for climate change. IPCC (Intergov. Panel Clim. Change), [2018](#page-87-1) and Owusu and Asumadu-Sarkodie, [2016](#page-88-0) define Carbon dioxide  $(CO_2)$ , Methane  $(CH_4)$ , Nitrous oxide  $(N_2O)$ , Hydrofluorocarbons (HFCs), Perfluorocarbons (PFCs) and Sulphur hexafluoride ( $SF_6$ ) as GHGs. The GHGs block the radiation from earth towards space. Thus more radiation stays within the atmosphere and raises the temperature on earth (El Zein and Chehayeb, [2015](#page-86-1)).

 $CO<sub>2</sub>$  has the greatest impact on the ozone layer. It has increased by more than 35% since the pre-industrial time (Reay et al., [2007](#page-88-1); Owusu and

#### 1. Introduction

Asumadu-Sarkodie, [2016](#page-88-0)). As Reay et al., [2007](#page-88-1) shows, the two main sectors of  $CO<sub>2</sub>$ emission are energy production and industrial processes. Since these two sectors are the biggest  $CO<sub>2</sub>$  producers, small reductions in these sectors have a massive impact on overall  $CO<sub>2</sub>$  emissions. Nowadays, the industry grows nearly every year, with two exceptions in the past twenty years. The first exception was the economic crisis in 2008 and the second exception is the current corona crisis. The crises had a massive impact on the industry and the gross domestic product (GDP). The GDP is an indicator of the development in industry and economy. Over the past ten years, the GDP grows by 0.7% every year on average in Austria (WKO Austria, [2020](#page-89-2)). A growing industry leads to increased energy consumption. Flauger et al., [2020](#page-86-2) say that the need for energy will rise tremendously in the next years. They see one of the main reasons for the increasing consumption in the electrification of industrial processes. Instead of using coal, oil, and gas for heating, cooling, and energy-intensive processes, the industry uses electric energy to generate heat for processes. Currently, the primary sources of electric energy are coal, oil, and gas (International Energy Agency, [2019](#page-87-2)), but electricity can also be created from renewable sources like water, wind, or solar power. Then electric energy is renewable and  $CO<sub>2</sub>$ -neutral energy. Forecasts say that the rise of renewable energy resources cannot cover the demand of energy needed in the next years. The world cannot achieve 100% renewable energy without (i) a decline in energy consumption, (ii) improvement of the efficiency in energy generation, and (iii) the increase of the share of renewable energy (Masson-Delmotte et al., [2018](#page-88-2)).

The growing industry does not only mean growing energy consumption. It also implies modernization, which newer plants reach. In contrast to earlier plants, newer ones use a massive number of sensors to collect data. The data is meaningful for the product's quality, process monitoring, energy supply monitoring, controlling, and various other parameters. The gathered data controls the machines' process and allows a fine-grained setting because the controller can react faster with the permanently added data. The permanent measurements and the resulting settings lead to products of higher quality and a lower rejection rate.

Digitization and the continuous increase of computational power are fundamental parts of Industry 4.0 (Lasi et al., [2014](#page-87-3); Bendel, [2019](#page-86-3)). In Industry 4.0, machines and production equipment are capable of decision-making. With the data from sensors, connections to other devices, and Artificial intelligence (AI), which processed the data, it is possible to autonomously monitor the system and support operators and decision-makers in defining the next steps. With continuous learning, AI can more and more make its own decisions.

System control and control optimization are important features in Industry 4.0 (Uhlemann, Lehmann, and Steinhilper, [2017](#page-89-3)). The first step requires a digital copy of the system. With the collected data, the digital copy can emulate the represented system, and in the last step, the digital system can control the real system. This system is called a Digital Twin (DT). For more details on DT, see Chapter [2](#page-12-1).1. With a DT, it is possible to perform optimization on a digital level and simulate its new behavior. With this setup, it is easy to optimize the system for different parameters without affecting the real production system. One of these parameters is energy consumption and, in conclusion, the reduction of  $CO<sub>2</sub>$ emissions.

There are multiple ways to reduce CO<sub>2</sub> emissions. On the one hand, there is renewable energy, and on the other hand, there is energy optimization. Since the industry is strongly dependent on the environmental conditions, it is not always possible to rely on renewable energy, but reducing energy consumption by energy optimization is still a possibility. The optimization needs monitoring, screening, and data analysis to retrieve knowledge from the massive amount of data generated by sensors and measuring systems. The data elaboration is part of Industry 4.0. The amount of data also allows the creation of DTs. DTs are a key-enabler for optimization and, therefore, even an introductory module in  $CO<sub>2</sub>$  reduction.

This thesis is part of the Austrian Research Promotion Agency (FFG) project DigitalEnergyTwin in cooperation with AT&S Austria as industry partner. It is part of a vast project and tries to improve, simplify, and extend existing parts in a DT project. The thesis has two main parts. One part is the DT modeling of an electroplating plant, and the second part is data prediction as input for the DT model. The model was developed in cooperation with project partners within the FFG project. AEE INTEC provides a model of the general structure of the electroplating plant. This thesis adds the control logic for the machines, various processes, and safety measures in the plant. The prediction part is split up into two subparts. The first prediction part is

#### 1. Introduction

the electroplating process. In this part, the electroplating process from the electroplating plant is recreated in a lab experiment. During the experiment, data about the process is collected. With the gathered data, various machine learning algorithms are trained and tested. The primary purposes are to understand the electroplating process, collect data for data models, and develop them. The output of the models is then input for the DT model. This part is a collaboration with Michael Grömer from the physics department at TU Graz. The second prediction part is about the approximation of sensor data. The plant operators collect a massive amount of sensor data from their plants. This amount of data is too big, so it is impossible to process this in the DT model. A lightweight alternative is needed to get results without loading all the data at once. The result is an algorithm to approximate the outcome. Further, it also allows predicting data in the future.

The thesis is divided into five chapters. Chapter [2](#page-12-0) describes the basics of used methods and implementations. It furthermore links the independent parts together to create an overview of the work. The chapter is about Digital Twins, the modeling language Modelica, and its user interface Dymola. Further, it holds information about electroplating, various data models, and approximation methods. Afterwards, Chapter [3](#page-34-0) presents the evaluation and the individual components' results. Chapter  $\frac{1}{4}$  $\frac{1}{4}$  $\frac{1}{4}$  is about the current status of the project and how the thesis results work are included. It also contains personal assessments and interpretations. Chapter [5](#page-82-0) completes the work and gives an outlook on further work and upcoming issues based on the thesis. [A](#page-100-0)ppendix  $\overline{A}$  contains the overview of energy flows which is the basis for the DT model. For the sake of completeness and traceability, Appendix [B](#page-102-0) and Appendix [C](#page-122-0) hold the handwritten and digitized data of the lab experiments.

# <span id="page-12-0"></span>2. Methods

## <span id="page-12-1"></span>2.1. Digital Energy Twin

### <span id="page-12-2"></span>2.1.1. Digital Twin

*A Digital Twin is an integrated multiphysics, multiscale, probabilistic simulation of an as-built vehicle or system that uses the best available physical models, sensor updates, fleet history, etc., to mirror the life of its corresponding [. . . ] twin.* (Glaessgen and Stargel, [2012](#page-87-4))

A Digital Twin (DT) uses a real-world object as an example and mimics it. The first definition was given by Grieves, [2002](#page-87-5) during a presentation for the product lifecycle management and was then defined by Tao et al., [2018](#page-89-4). It is a digital counterpart to the physical object. Instead of a controller in the physical object, algorithms replace the control logic. The algorithms mirror the behavior. The mirroring uses an "as best as possible" approach, where the algorithms are optimized to act like they can replace the realworld controller. The use of digital twins can have different purposes. A primary goal is the accurate and precise representation of the behavior of the physical object. A further objective is to decouple the digital object from the real-world object, do experiments on it, and see the influence of various changes on the system's different parameters without bothering the real-world processes. The experiments can have multiple purposes of use. Optimization is one of them. The main targets for optimizations are energy optimizations, duration reduction, lower costs, or  $CO<sub>2</sub>$  reduction. DT allows investigating different control strategies that minimize energy consumption (thus  $CO<sub>2</sub>$ ). Duration improvement reduces the needed time to produce output with consistent quality (Gaul, [2018](#page-86-4)).

#### 2. Methods

Kritzinger et al., [2018](#page-87-6) has created an overview of different papers aiming at the DT and gives a classification for different categories of DTs. Figure [2](#page-14-0).1 shows these classifications. The DT has several classes. The basic class is the Digital Model (DM). The DM has a physical object and a digital object, which represents the physical object. Data exchange between the two objects works manually. The user both extracts the physical object's data and after that inserts it into the digital object by hand. It has the advantage that the user can check inserted data beforehand, so no damage to the physical object happens. Checking data from the physical objects also has the advantage that the user gets an overview of the data and detects failed measurement. The user can use the knowledge to improve the digital object and make it more fault tolerant. This is the first stage in the development process. It helps to improve the safety of the digital object cleverly. A Digital Shadow (DS) is a more advanced model. There is also a physical object and a digital object in the DS, but data transfers are different. The data from the physical object to the digital object works automatically. Thus, the objects have a common interface which they use to transfer data to the digital object. Various patterns fulfill automatic data transfer. A simple approach is a file-based transfer, where the physical object writes to a file, and the digital object reads the file and processes the data. A more advanced approach is a sender and receiver pattern where the physical object sends data, and the digital object receives it. The digital object returns the processed data manually to the physical object. Thus, the digital object computes suggestions, depending on the physical object's data, but the decision of implementing them is up to the operator. The most advanced class in this classification is the Digital Twin (DT). This class transfers all data automatically. In extension to the DS, the DT also has automatic data transfer from the digital object to the physical object. It means that the digital object is aware of control for the physical object. The computed data controls the physical object actively. With a DT, the digital object has huge power over the physical object.

<span id="page-14-0"></span>

Figure 2.1.: Digital Twin vs. Digital Shadow vs. Digital Model (Kritzinger et al., [2018](#page-87-6)) Digitization has a physical object as a role model. The digital object tries to reproduce the behavior of the physical object. Depending on the way of data transfer, it is possible to distinguish between 3 types:

> In a **Digital Model**, the data transfer is completely manual. Data from the physical object are transferred by hand to the digital object and vice versa. With a **Digital Shadow**, the transfer from the physical object to the digital object is automated. Still, the transfer back to the physical object has to be done by hand. A **Digital Twin** automates every connection between the physical and the digital object.

### <span id="page-15-0"></span>2.1.2. From Digital Twin to Digital Energy Twin

A DT represents a physical object in various ways. A special derivate is the Digital Energy Twin (DET). As the name already says, the main target of the DET is energy. First, it simulates the energy consumption of a physical object with given data of any kind. The data can be energy consumptions or any process data, enabling energy consumption calculation or prediction.

In the course of this thesis, the electroplating process is of special importance. The computations of the energy consumption for an electroplating plant need current, voltage, and plating duration. Computing energy consumption of pumps needs information about workload, degree of efficiency, and a function mapping to energy consumption. Secondly, the comparison between the plant's calculated energy consumption and the real measured energy consumption happens. When there are several measuring points, the process is more straightforward because a divide and conquer approach is used by dividing the whole energy consumption into smaller chunks and compare these chunks with the calculated one. This approach finds errors in the calculations more quickly.

## <span id="page-16-0"></span>2.2. Modelica and Dymola

This section is about Modelica, Dymola, and the implementation of an electroplating plant using these tools. The combination of Chapter [2](#page-12-1).1 and Chapter [2](#page-16-0).2 create a functional simulation of the plant. For results, see Chapter [3](#page-34-1).1.

### <span id="page-16-1"></span>2.2.1. Modelica

Modelica (Ramírez et al., [2017](#page-88-3)) is an open-source, object-oriented modeling language for physical systems. Modelica compiles models into C-code. Thus, models are computed efficiently on the CPU. Modelica uses acausal modeling; this means that the system is described via differential-algebraic equations (DAE). A equation has the form  $F(\dot{x}(t), x(t), y(t)) = 0$ . With given conditions, Modelica solves the equations for unknowns. This acausal approach supports the user by focusing on the modeling part rather than on Modelica's background. Variables have a special syntax, which allows adding further information to them. The listing below shows a definition of a parameter with details describing the unit of the parameter and an explanation of the parameter.

parameter Real duration(unit="s")"Duration time in process X";

The unit system enables Modelica to evaluate given equations for correctness. By default, Modelica supports a vast set of physical models. The standard library contains elements to model mechanical, electrical, thermal, fluid, and many other physical systems (Modelica Association, [2020](#page-88-4)). Furthermore, a lot of free libraries provided by the community are available. This project mainly uses the opensource library *Buildings* (Wetter et al., [2014](#page-89-5)). The library offers great support for every kind of heating, cooling, and general energy supplies in any existing kind.

### <span id="page-17-0"></span>2.2.2. Dymola

Modelica has multiple graphical user interfaces (GUI). Widely used GUIs are Dymola<sup>[1](#page-17-1)</sup>, SimulationX<sup>[2](#page-17-2)</sup>, Wolfram SystemModeler<sup>[3](#page-17-3)</sup>, and Openmodelica<sup>[4](#page-17-4)</sup>. This project uses Dymola as its GUI. Peter Fritzson, [2011](#page-86-5) and Fritzson and Thiele, [2016](#page-86-6) explain Modelica and present details about the langauge. Modelica is text-based. The GUI adds graphical elements for coding. The user can place blocks on the program surface. The blocks act as placeholders for classes. Wires connect blocks in a bi-directional way. The connectors are inputs and outputs simultaneously, which is useful for physical models when for example, two containers connected via a pipe exchange a fluid. When the fluid flows from container *A* to container *B*, the connector at container *A* works as output, and the connector at container *B* works as input. Vice versa, when the fluid flows from container *B* to container *A*, *B*'s connector is the output, and *A*'s connector works as input. The GUI also has an interface for simulation. This section shows the simulation results in graphs and can display the results of a simulation. Figure [2](#page-18-0).2 shows a screenshot of the Dymola GUI with a loaded example program.

<span id="page-17-3"></span><sup>3</sup><https://www.wolfram.com/system-modeler/>

<span id="page-17-2"></span><span id="page-17-1"></span><sup>1</sup><https://www.3ds.com/de/produkte-und-services/catia/produkte/dymola/> <sup>2</sup><https://www.simulationx.de/>

<span id="page-17-4"></span><sup>4</sup><https://www.openmodelica.org/>

<span id="page-18-0"></span>

#### Figure 2.2.: A screenshot of the Dymola GUI

On the left side, the GUI shows the package browser, where all available packages and modules are listed. The main windows show a heating system model. In the right bottom corner is the switch between the modeling interface and the simulation interface.

## <span id="page-19-0"></span>2.3. Electroplating

Electroplating (Jelinek, [2013](#page-87-7); Kanani, [2009](#page-87-8); Unruh, [2016](#page-89-6)) is the electrochemical process of deposing metal ions on a conductive material. Figure [2](#page-20-0).3 shows a basic setup for electroplating. The main components are an energy source, a cathode, an anode, and an electrolyte. The energy source must be direct current (DC). Alternating current (AC) would reverse the ongoing electrochemical process twice per swinging period. The anode consists either of the material which plates the cathode or of inert material. Inert material does not dissolve during the electroplating process. It is just a tool to close the electric circuit. In this case, the dissolved metal ions in the electrolyte plate the cathode. The concentration of deposing metal ions in the electrolyte determines the maximum plating thickness. When the anode is a non-inert material, it dissolves during the process. In this case, the anode and the electrolyte must match. The electrolyte must have the same ions dissolved as the anode delivers when it dissolves. When the anode is copper, the electrolyte also must be made of a copper ion. Thus a copper(II) sulfate  $(CuSO<sub>4</sub>)$  solution suits well. One would use silver nitrate  $(AgNO<sub>3</sub>)$  as an electrolyte for a silver anode. The electrolyte's main purpose is the transportation of ions to the cathode, where they got deposited. When there is a dissolvable anode, the concentration of metal  $(ME^{+/2+})$  ions stays constant during the process. The electrolyte creates a concentration gradient from the anode to the cathode because the anode adds ions to the solution, and the cathode removes them. Mixing the liquid reduces this counterproductive effect. Metal ions deposit on the cathode. The process plates every part of the cathode's surface, which is in touch with the electrolyte. The plating's quality measurements are the surface quality, the plating thickness, and the layer's ionic structure. Electric field lines, the current density, process duration, and temperature are responsible for the plating quality. The temperature influences the ionic structure of the electroplated layer. It does not have an impact on the plating thickness. Current density and the process duration mainly define the plating thickness. Straight field lines are essential for an even overall layer thickness.

<span id="page-20-0"></span>

Figure 2.3.: Schema electroplating

Electroplating has four main parts. The energy source delivers direct current electricity for the reaction. The anode dissolves or is inert to the electrochemical process, while on the cathode, the material deposits. The electrolyte dissolves the metal ions and is responsible for the ion transport to the cathode, where the ions solidify. Adapted from Unruh, [2016](#page-89-6).

<span id="page-21-0"></span>

Figure 2.4.: Field lines experiment

The experiment shows field lines in a magnetic ambient. The experiment uses a magnet with north-pole (-) and south-pole (+) and iron filings. The iron filings are magnetized and align themselves according to the field lines. Figure [2](#page-22-0).5 vividly illustrates the result of the experiment. (Black and Davis, [1922](#page-86-7))

<span id="page-22-0"></span>

#### Figure 2.5.: Electric field lines schema

Electric field lines emerge between a positive and a negative electric charge and represent the Coulomb force. They always flow from the positive to the negative pole in a directed way. Electroplating needs as many straight field lines as possible. Curved field lines generate edge cases where more material deposits and result in an uneven layer. Adapted from Black and Davis, [1922](#page-86-7); Unruh, [2016](#page-89-6).

#### 2. Methods

Figure [2](#page-21-0).4 shows field lines in an experiment. The iron filings align themselves according to the field lines. The curvature of the field lines leads to an uneven distribution of the iron filings on the poles' edges. The same effect also takes place in electroplating. In areas where many field lines hit, more material deposits than in areas where fewer field lines hit. Figure [2](#page-22-0).5 shows a comparison between straight and curved field lines. The upper case represents the experiment in Figure [2](#page-21-0).4. The lower case is an ideal setup for an even deposition of material on a workpiece. The lab experiment (see Chapter [2](#page-24-0).3.1) and the industrial process aim to have straight field lines. Both cannot use the ideal setup because of various reasons. On the one hand, a workpiece does not always have a cylindric shape, and on the other hand, it is way more expensive to build the anode as a cylinder and fill it with electrolyte. It is also hard to transfer the workpiece through the cylindric anode in an automated, continuous way.

The base of electroplating are redox reactions. Redox reactions are a combination of reduction reactions and oxidation reactions. Redox reactions include the movement of electrons (e – ) from the reduction reaction to the oxidation reaction. In electroplating, the reduction reaction takes place on the cathode. Reaction  $R_1$  $R_1$  shows the general reaction of deposing a metal ion.  $X \in \mathbb{N}$  is the number of charges. For metal ions, it is usually 1 or 2. For example copper is a doubly positively charged ion  $Cu^{2+}$ , here  $X = 2$ .

<span id="page-23-0"></span>
$$
Me^{X+} + X \cdot e^- \longrightarrow Me
$$
 [R1]

On the anode, the oxidation reaction takes place. Reaction R[2](#page-23-1) shows the general reaction of dissolving metal.

<span id="page-23-1"></span>
$$
Me \longrightarrow Me^{X+} + X \cdot e^{-} \qquad [R2]
$$

In sum, both reactions take place at the same time. The redox reaction is the resulting reaction. Reaction  $R_3$  $R_3$  shows the whole redox reaction.

<span id="page-23-2"></span>
$$
Me \longrightarrow Me
$$
 [R3]

The reactions do not work automatically, they need an additional energy source. Figure [2](#page-20-0).3 shows the energy source. Electroplating uses a direct current energy source, which transfers the electrons from the anode to the cathode. In other words: Transfering the electrons from the oxidation

reaction to the reduction reaction. The plus pole of the energy source is an electron sink, and the minus pole is an electron source. This explains why the layer thickness depends on the time and the Amperage. The longer the process lasts, the longer electrons can move, and the higher the amperage, the more electrons move in a specific time interval.

### <span id="page-24-0"></span>2.3.1. Lab-scale experiment

The lab experiment aims to understand the electroplating process, the influence of various parameters on the layer thickness, and the layer quality. With the gathered data, a data model is created to describe the electroplating process. All these steps help to better understand the process on an industrial scale. The experiment took place at the department of experimental physics at TU Graz. Within three weeks, over 200 electroplating experiments took place. The experiment uses copper as anode and as cathode. Resulting, the electrolyte is a  $CuSO<sub>4</sub>$  solution. The energy source is a laboratory power supply. To keep a constant temperature, water baths are used.

The varied parameters are duration time (*t*), temperature (*T*), and current density (*j*). Duration time and current density directly impact the coating thickness, while temperature does not directly impact the coating thickness. Temperature influences the structure of the crystal lattice. Its results are smoother and trouble-free surfaces and layers. Equation [2](#page-24-1).1 (Unruh, [2016](#page-89-6)) shows the influence of the parameters on the coating thickness and Equation [2](#page-25-2).2 explains the term current density.

<span id="page-24-1"></span>
$$
t_e = \frac{Ae}{\varrho} * \vec{j} * t * \eta_i \tag{2.1}
$$

where:

- $t_e$  = Coating thickness
- *Ae* = Electrochemical equivalent, *Ae*(Cu) = 1.19 g A<sup>-1</sup> h<sup>-1</sup>
- $\varrho$  = Density of the layer material,  $\varrho$ (*Cu*) = 8.96 g cm<sup>-3</sup>
- $\vec{j}$  = Current density, see Equation [2](#page-25-2).2
- $t =$ Duration time
- $\eta_i$  = Current efficiency,  $\eta_i$ (*Cu*) ~ 0.97

$$
\vec{j} = \frac{I}{A} \tag{2.2}
$$

<span id="page-25-2"></span>where:

 $\vec{j}$  = Current density  $I =$ Current

 $A =$ Coating area

## <span id="page-25-0"></span>2.4. Data models

### <span id="page-25-1"></span>2.4.1. Machine learning

Machine learning (ML) (Aggarwal, [2020](#page-86-8)) is a method to generate knowledge from existing data. During training, machine learning algorithms learn parameters for statistic models using training data. After the training, the ML algorithm can predict further unknown data and does not need the training data anymore. During the training phase, the algorithm finds patterns and correlations in the given data. The project uses supervised learning for the training phase. It means that the training data consists of input values and the corresponding output value (= result), and the algorithm tries to fit the data in the best possible way. A problem of ML is overfitting. Figure [2](#page-26-0).6 shows an example of overfitting. Here, a linear and a polynomial model approximate a noisy linear function. While the polynomial function has excellent results on the training set by overfitting the values, it fails on the test data set. There are various models for ML. The next sections give an overview of the used models in this thesis.

#### Neural Networks

*The neural network is this kind of technology that is not an algorithm, it is a network that has weights on it, and you can adjust the weights so that it learns. You teach it through trials.* – Howard Rheingold (M. B. Patel, J. N. Patel, and Bhilota, [2020](#page-88-5))

<span id="page-26-0"></span>

#### Figure 2.6.: Overfitting

The data (blue and orange dots) shows a noisy linear function. The blue dots represent the training data set, and the orange dots represent the test set. With the training set, two models are trained. The green line represents the linear model, and the red line represents a polynomial model. The linear model has a *R* <sup>2</sup> of 0.96 on the training data, while the polynomial model fits the data perfectly and has a  $R^2$  of 1. But when the models apply to the test data, the  $R^2$ differ a lot. The  $R^2$  of the linear model is 0.99, while the  $R^2$  of the polynomial model drops to 0.84. It is a typical case of overfitting. The model fits the training data without having the background of the data in mind, and so it fails on further data.

#### 2. Methods

The idea of Neural Networks (NN)(Rojas, [1996](#page-88-6); Rashid, [2017](#page-88-7); Yiu, [2019](#page-89-7)a; Sai Ajay, [2020](#page-89-8)) is based on the way the human brain works. Neurons send signals when they are triggered. But instead of chemical reactions triggering the signal, mathematical equations are the reason for the reaction.

A NN has an input layer, an output layer and arbitrary many hidden layers. The input layer takes the input to the NN. The inputs are called features. A feature can be a measurement when we want to do regression, or it can be a pixel of an image for classification. The output layer represents the result of the prediction from the NN. In regression, the output layer consists of one neuron, which holds the result. In classification, the output layer has *n* neurons, where *n* is the number of classes. Each of the *n* neurons then holds the probability that the input matches the represented class. Each neuron in a layer is connected with each neuron in the next and the previous layer. Each connection has a weight, which is multiplied by the value of the outgoing neuron before it comes to the incoming neuron. A neuron holds a bias, which is added to the incoming values, and the result is then the input parameter for a non-linear function that produces the output of the neuron. The non-linear function can, for example, be a sigmoid function.

The training of the NN is based on forward propagation and backward propagation. During the learning phase, the features traverse the NN, and the output is compared to the set's label using a cost function. Adapting the weights and biases can minimize the cost function. So backward propagation uses the result of the cost function to traverse the neural network backward from the output to the input. In the backward propagation, the algorithm adapts the weights and biases to minimize the cost function. For the implementation, scikit-learn (Pedregosa et al., [2011](#page-88-8)) as one of the most popular and established Python packages is used.

#### Linear regression

Linear regression (LR) (Rohith, [2018](#page-88-9)) is a statistical method used in ML. The main goal is to fit given data with a linear function. The primary condition is to minimize the mean square error (MSE) between the predicted value and the real data. The LR uses a cost function and minimizes it. The cost

function  $(=MSE)$ 

$$
J = \frac{1}{n} \sum_{i=1}^{n} (pred_i - y_i)^2
$$
 (2.3)

where:

 $n =$  Number of data points  $pred_i$  = Predicted value, see Equation [2](#page-28-0).4  $y_i$  = Data point

describes the error between the prediction and the real data. The predicted value (*pred<sup>i</sup>* ) is calculated by a linear function.

<span id="page-28-0"></span>
$$
pred_i = d + \sum_{j=1}^{m} w_j * x_j \tag{2.4}
$$

where:

 $w =$  Weight of the parameter  $x =$ Dependend variables  $d =$ Initial value

The LR uses the equations and finds the best fitting parameters  $(w_j, d)$  to minimize the MSE. For this, it uses a gradient descend method.

#### Polynomial regression

Polynomial regression (PR) (Nocedal, [2006](#page-88-10)) uses the same methods as LR, but instead of a linear function, it uses a polynomial function as a cost function. The degree of the polynomial is a choosable parameter. The choice of the degree determines whether the calculated function overfits or not. Choosing the degree in about the number of variables leads to a well-fitting polynomial. If the degree is 1, the polynomial regression becomes a linear regression.

#### 2. Methods

#### Decision tree regression

Decision tree regression (DTR) (Li, [2019](#page-87-9)) is also a supervised machine learning algorithm. During the training phase, the algorithm builds a binary tree with the decision results in the leafs. Each node has a true/false question, which is answered with the given data to create a prediction. The information gained by the split determines the splitting of the data in the left or right subtree. The algorithm maximizes the information gain (IG) for each split. For this, it optimizes the objective function.

$$
IG(D_p, f) = I(D_p) - \left(\frac{N_l}{N_p}I(D_l) + \frac{N_r}{N_p}I(D_r)\right)
$$
 (2.5)

where:

 $IG =$ Information gain  $I =$  Impurity function  $f =$ The feature for the node's criterion  $D_p =$  Dataset in the parent node  $D_l$  = Dataset in the left child node  $D_r =$  Dataset in the right child node  $N_p$  = Number of data points in the parent node  $N_l$  = Number of data points in the left child node  $N_r$  = Number of data points in the right child node

This means that the higher the parent node's impurity and the smaller the child node's impurity, the higher the information gain. In regression, the impurity function is the MSE between the real and the predicted value, as it was already in LR and PR. The DTR regression also gives information about the importance of the features. Figure [2](#page-30-0).7 shows a decision tree of depth 2 for the lab experiment. It shows that the most important features are the current density and duration. The DTR is likely to overfit, so the tree's maximal depth is an important parameter to reduce this error. Figure [2](#page-31-0).8 shows the result of the decision tree from Figure [2](#page-30-0).7. All predictions are in one of the four possible categories. Deeper trees result in more categories and subsequently also in more precise predictions. The bucket behavior of the DTRs leads to a further disadvantage. It is impossible to detect correlations

between the data; the model can only make new predictions within the training data set. Outliers are always assigned to the nearest bucket.

<span id="page-30-0"></span>

Figure 2.7.: Decision tree

The tree has a maximal depth of 2. In each node, the first line shows the condition, the second line shows the calculated MSE, and the third line shows the number of samples. The last line is the corresponding value.

#### Random forest regression

Random forest regression (RFR) (Yiu, [2019](#page-89-9)b) is a successor of DTR. It uses many uncorrelated DTR trees to vote for a result. The significant advantage is that this algorithm uses the wisdom of the crowd as a key feature. It means that there are many votes for the result, and when there is a maleficent vote, it is overruled by many other correct votes. The single DTR trees are uncorrelated. This ensures the statistical correctness of the algorithm. Two mechanisms grant it. The first one is Bootstrap Aggregation (Bagging). Small changes in training data lead to a totally different tree. When the training data consists of *N* data points, each tree chooses a random set of the training data. The newly selected data set has a size of *N* again; it also means that some of the new data sets contain the same data points multiple times. With many different training sets, many varied decision trees are generated. The second mechanism also builds on the sensitivity of the trees. Feature randomness lets each of the trees choose features randomly for all

<span id="page-31-0"></span>

#### Decision Tree (max\_depth =  $2$ )

#### Figure 2.8.: DTR data and predictions

The black dots represent the training data. The black crosses stand for the test data. Training data trains a DTR model of depth 2. The red dots represent the predictions of the model. The  $R^2$  is 0.72 on the training set and 0.59 on the test set. The model is not very accurate, but it shows the concept of DTRs quite well. Deeper DTRs are more accurate.

existing features. So various combinations of features a tree can choose from are created.

## <span id="page-32-0"></span>2.5. Function-based approximation of data

During the project, the partners had recorded an enormous amount of data, which is used in the DT models. Most of the data describe temperature curves, voltage, and amperage values. All the values follow a specific pattern influenced by the laws of physics and the plant's control strategy. For example, the temperature curves are nearly constant during production. They only vary slightly around the set temperature. When the plant gets turned off, the temperature falls according to the cooling curve of fluids

$$
T(t) = T_A + (T_0 - T_A) * e^{-c_c t}.
$$
 (2.6)

where:

- $T(t)$  = Temperature at a given time *t*
- $t =$ The time to calculate the temperature
- $T_A$  = The temperature of the ambient
- $T_0$  = The temperature of the fluid at the beginning of the cooldown
- $c_c = A$  coefficient affected by the volume of the fluid, the surface,<br> $c_c = Hermal$  capacity and the density of the fluid
	- thermal capacity, and the density of the fluid

When production starts again, a linear function describes a head up

$$
T(t) = T_0 + c_h t. \tag{2.7}
$$

where:

- $T(t)$  = Temperature at a given time *t*
- $t =$ The time to calculate the temperature
- $T_0$  = The temperature of the fluid at the beginning of the head up
- A coefficient affected by the volume of the fluid, the surface,
- $c_h$  =  $\frac{1}{2}$  determinent and the density of the fluid

#### 2. Methods

Using the given data for further processing leads to a massive computing overhead. The shown equations serve the same purpose but have a lot less computing overhead. The collected data are still in use. With the data, the functions are parametrized. The parametrization is a one-time overhead that can be reused. For results see Chapter [3](#page-69-0).3.

# <span id="page-34-0"></span>3. Results

## <span id="page-34-1"></span>3.1. Digital Energy Twin Model

## <span id="page-34-2"></span>3.1.1. The Digital Twin of an electroplating plant

This Section represents the results of Chapter [2](#page-12-1).1 and Chapter [2](#page-16-0).2. The model is based on the scheme from Appendix [A](#page-100-0) (ENERTEC Naftz & Partner GmbH & Co KG, [2020](#page-86-9)). Figure [3](#page-35-0).1 shows the created model in Dymola. The energy scheme describes the flow of energy in the system. The scheme has two parts. On the left-hand side, it shows the first plant (Werk 1), and on the right-hand side, it shows the second plant (Werk 2). This thesis deals only with the left-hand side because it is the central part of its development. The colors of the connections stand for the pipes' temperatures, where red connections mean hot fluid and blue connections mean cold fluid.

The central parts of the scheme are the two tanks. One contains cold water (KW Becken Werk 1), and the other contains warm water (WW Becken Werk 1). Whenever the warm water level rises too high, the water dumps into the nearby river (End-Kontrolle KW 1). The cold water tank stores water for the process. When the water level gets too low, it adds water from a well (Brunnenwasser) and tempers it with water from the warm water tank. The cold water tank delivers the water for process cooling, and the warm water tank provides rinsing water. A hydraulic separator generates hot water. It mainly uses waste heat from the other machines for the heat up. If not enough waste heat is present, the plant has a few options to generate heat. It can activate a heat pump (WP) and cool down warm water from the warm water tank and bring it to the cold water tank. In the last instance, natural gas is used to heat the hydraulic separator.

<span id="page-35-0"></span>

Figure 3.1.: Energy flow model in Modelica
The Modelica model in Figure [3](#page-35-0).1 aims to imitate the energy scheme, the control of the machines, and the security mechanisms. It also uses the collected data from the real plant for the simulations. One of the main advantages of Modelica is visualization. After studying the energy scheme, the Modelica model looks similar to the real plant, and it is easy to navigate in the model and find the components. The next part shows the implemented control mechanisms. All implementations are designed as state machines to avoid chattering. Chattering is a misbehavior in numerical methods. At a specific point, the calculation jumps between multiple states within milliseconds. A specific point can be a threshold determining on or off, for example. State machines also use delays before they change to the next state. That avoids chattering and is also used to simulate the start-up time and the switch-off time.

## Security mechanism - Overflow detection

<span id="page-37-0"></span>



On the left-hand side, the model shows the inputs. The dark blue painted arrows are the inputs and hold values for the current water level and the maximum water level in the tank. The model has three thresholds (Overflow, Normal, Empty) determining the four possible levels. Logical equations transform the input values and the thresholds into four boolean signals for the state machine in the middle of the picture. Depending on the state machine's boolean input, the state, and the delays, the state machine can produce an overflow alert. The alert is represented by a boolean signal and stored in a flip-flop register. The right-hand side shows the outputs of the overflow detection.

Overflow detection is a safety mechanism. It controls the water level in the tanks. When the level rises too high, it directs water into the nearby river. Figure [3](#page-37-0).2 shows the overflow detection. The core component is the state machine in the middle of the model. It has six states. The initial state is the starting point for the calculations. After the initial state, the machine has four alternating states. In each iteration, only one of the four possible states is active. By definition of the module, the first state that gets activated is the prior one used in the current iteration. The four alternating states all have a separate boolean output. It determines whether the output shows

an overflow alert or not. The last state is an end state. It does not have any effect. It is needed for the completeness of the program. Each state has a transition block (black vertical bar) in front and behind it. Transitions serve two purposes. At first, they release or block a path depending on the boolean input they receive. If there is no input, they always release the path. The second purpose is to introduce delays. As otherwise, the chattering effect would happen. From the data around the water tanks, it is clear that an overflow takes several minutes to happen, and it is not necessary to react to an overflow within seconds. It is sufficient to respond within a few minutes. So the transition between the last state and the initial state has a delay of five minutes. The following equations describe the input for the transitions.

$$
\sigma_{\text{FULL}} = l_{\text{curr}} \ge (l_{\text{max}} * t_{\text{FULL}}) \tag{3.1}
$$

$$
\sigma_{A. FULL} = (l_{curr} < (l_{max} * t_{FULL})) \land (l_{curr} \geq (l_{max} * t_{NORMAL})) \tag{3.2}
$$

$$
\sigma_{A.EMPTY} = (l_{curr} < (l_{max} * t_{NORMAL})) \land (l_{curr} \geq (l_{max} * t_{EMPTY})) \tag{3.3}
$$

$$
\sigma_{EMPTY} = l_{curr} < (l_{max} * t_{EMPTY}) \tag{3.4}
$$

where:



A flip-flop register stores the result of the overflow detection. For the flip-flop register following equations hold.

$$
S = s_{FULL} \tag{3.5}
$$

$$
R = s_{A, FULL} \vee s_{A,EMPTY} \vee s_{EMPTY}
$$
 (3.6)

where:



<span id="page-39-0"></span>



The upper figure shows the current, maximum, and trigger level of the overflow detection. Every time the current level goes above the trigger level (at hour 644 and hour 662), the lower figure switches from off to on. When the level falls below the trigger level (at hour 648 and hour 664), also the signal turns off.

Figure [3](#page-39-0).3 shows the result of the overflow detection. The overflow detection works as expected. Whenever the water level rises above the trigger level, the signal turns on, and then the water level drops again. Modelica models crash if there is an overflow in a component. This security mechanism saves the model from crashing and also imitates the real object well.

<span id="page-40-0"></span>

### Process mechanism - Compressor control

#### Figure 3.4.: Compressor control model

The model only has one input, which is the current air consumption of the plant. There are also parameters for the power of the compressors. The state machine in the middle of the figure has two states. One represents the first production mode, where two dynamic compressors handle the compressed air production. The second state takes the case when the two dynamic compressors do not have enough power to handle the required air, and the third compressor needs to be activated. Depending on the state, the model calculates the dynamic compressors' power level and the static compressor's control signal.

The plant uses compressed air to drive drilling machines. Depending on the capacity utilization of the drilling machines, the compressed air consumption varies. The plant uses three compressors to produce compressed air. Two compressors are dynamic compressors with dynamic rotation speed. The compressors are more energy-efficient than the third compressor, which is a static one. It can only be turned on or off. It is desired that the dynamic compressors run as much as possible, and that the static compressor is only active when the dynamic compressors cannot handle the current air consumption. Figure [3](#page-40-0).4 shows the compressor control model. It has an input for the current air consumption and uses maximal producible air amount as parameters. A central part is a state machine. In total, it has four states. An initial state, which is the entry for the model. Two states represent the production modes. The first mode handles the case that the dynamic compressors have enough power to produce the required air consumption. The second state assumes that all three compressors must be in operation to fulfill the compressed air needs. The program needs the last state for correctness reasons. To decide which state is active, the transitions (vertical black bars) have the following conditions.

$$
\sigma_{2C} = (\eta_A < 0.95 * (2 * \min(p_{C1}, p_{C2}))) \vee ((\eta_A < 0.75 * (2 * \min(p_{C1}, p_{C2}))) \wedge s_{2c}) \quad (3.7)
$$

$$
\sigma_{3C} = \neg(\eta_A < 0.95 * (2 * \min(p_{C1}, p_{C2})))
$$

(3.8)

where:



Depending on which transition is activated, a flip-flop register *R* stores the result. If *R* is one, then two compressors are active; otherwise, all three compressors are used. The flip-flop register directly influences the static compressor. If *R* is one, then the static compressor is off and if *R* is zero, then the static compressor is on. The power level for the dynamic compressors calculates as following.

$$
pl_{C1} = \begin{cases} \left(\frac{\eta_A}{2}\right) * p_{C1}^{-1} & \text{2 compression mode} \\ \left(\frac{\eta_A - p_S}{2}\right) * p_{C1}^{-1} & \text{3 compression mode} \end{cases}
$$
(3.9)

$$
pl_{C2} = \begin{cases} \left(\frac{\eta_A}{2}\right) * p_{C2}^{-1} & \text{2 compression mode} \\ \left(\frac{\eta_A - p_S}{2}\right) * p_{C2}^{-1} & \text{3 compression mode} \end{cases}
$$
 (3.10)

where:

 $pl_{C1}$  = Power level of compressor 1  $pl_{C2}$  = Power level of compressor 2  $\eta_A$  = The current compressed air consumption  $p_{C1}$  = Power of dynamic compressor 1  $p_{C2}$  = Power of dynamic compressor 2  $p<sub>S</sub>$  = Power of static compressor

Figure [3](#page-43-0).5 shows the result of the implemented algorithm. As the requirements request, the dynamic compressors run as much as possible. Only in peak times, the static compressor runs.

<span id="page-43-0"></span>



The upper figure shows the current air consumption. The data are randomly generated values from a test suit. The below figure shows the control signals for the compressors. The static compressor has a fixed rotational speed and has only two operating modes. It can be on or off. The dynamic compressors can vary the rotation speed. So the control signal is the share of the maximum power of the compressors. As long as the dynamic compressors can handle the air consumption, the static compressor is off (for example, between hours 34 - 50). But when the air consumption rises and the dynamic compressors reach their limits, the static compressor turns on and supports the compressed air production (e.g., hour 50).

<span id="page-44-0"></span>

## Process mechanism - Tank logic

## Figure 3.6.: Tank logic model

The tank logic module contains the logic for triggers depending on the tanks' water level and the water temperature. It includes the overflow logic from Figure [3](#page-37-0).2 as modules twice - one time for the cold water tank and one time for the warm water tank. Additionally, there are logical equations for heat pumps and heat exchangers switches. The module also decides if water from the well is needed and if the water in the cold water tank needs to heat up.

As shown in the overflow detection chapter, the tank levels are critical measurements for the control logic, but there is more information within the tank levels. The control state of other machines like heat exchangers or heat pumps depends on the tank levels. Figure [3](#page-44-0).6 shows the tank logic module, which controls several devices based on the tank levels. The inputs are temperatures and the water levels of the tanks. Additionally, boolean inputs determine the need for heat for the hydraulic separator and the low-temperature sector. As Appendix [A](#page-100-0) shows, there are many devices connected to the water tanks. A heat pump (WP) uses warm water to

generate extra heat for the hydraulic separator by cooling the warm water down and deliver it to the cold water tank. Another device is the heat exchanger (WT KW), which creates load for the chillers (KM 1-3), producing more waste heat, which heats the hydraulic separator. The last control unit is the refilling of the tank. There is also a control part for the inlet of well water and the cold water tank's tempering.

The following equations describe the state of the heat pump and the heat exchanger.

$$
r_{HP} = \eta_{HS} \wedge \lambda_{WW} \wedge \neg \omega_{KW} \tag{3.11}
$$

$$
r_{HE} = \eta_{LT} \wedge \lambda_{WW} \wedge \neg \omega_{KW} \tag{3.12}
$$

$$
\lambda_{WW} = s_{FULL} \vee s_{A.FULL} \tag{3.13}
$$

where:



The following equations describe the water level and the water temperature in the cold water tank.

$$
r_{FT} = \epsilon_{KW} \wedge \neg r_{HP} \wedge \neg r_{HE} \tag{3.14}
$$

$$
r_{HC} = \tau CW < t_{TEMP,LOW} \land \neg r_{HE} \land \neg r_{HP} \tag{3.15}
$$

where:



<span id="page-46-0"></span>

Figure 3.7.: Heat pump result

The heat pump is only active if the warm water level is high enough and the hydraulic separator needs the energy. The upper chart shows the warm water level and the control signal from the hydraulic separator. The lower graph shows the resulting control signal for the heat pump. The first on-phase (hour 14 to hour 18) of the heat pump shows the case where the hydraulic separator would need energy earlier, but there is too little water in the tank, and the controller cannot turn on the heat pump. After a while, the water level rises, and the heat pump turns on. The last on-phase shows the case where all conditions meet, and the heat pump is on.

<span id="page-47-0"></span>



The heat exchanger only works when the low-temperature sector needs coldness and the warm water level is high enough. The first on-phase shows the case where all conditions meet (hour 2 to hour 5). The second and the third on-phase show that the low-temperature sector needs more cold energy, but the water tank level is too low, so it is not always possible to provide.

<span id="page-48-0"></span>

## Figure 3.9.: Well inlet result

The well inlet is responsible for the refilling of the cold water tank. While the heat pump or the heat exchanger is running, it is not needed to use well water to refill the cold water tank because the two devices feed the cold water tank with water. If the cold water level is shallow, the well inlet is active no matter whether the heat exchanger or the heat pump runs. This the case for the last on-phase in the lower chart (hour 34 to hour 38). All the other cases do not have an urgent need, so the water feed from the heat pump and heat exchanger are sufficient. The upper chart shows the cold water level and the control signals for the heat pump and the heat exchanger's control signal.

<span id="page-49-0"></span>



If the cold water gets too cold, the warm water tank heats it if some conditions meet. The cold water tank must not have an overflow, and the warm water tank must contain enough water. Further, the heat exchanger and the heat pump should also be off, because otherwise they would produce tempered water. If the cold water temperature is far too low from production set temperature, it is acceptable to use the heat pump, the heat exchanger, and the warm water to heat the cold water tank. The first and the last on-phase in the lower chart show this behavior. The upper chart shows the necessary values for decision-making.

Figure [3](#page-46-0).7, Figure [3](#page-47-0).8,Figure [3](#page-48-0).9, and Figure 3.[10](#page-49-0) show the results of the control strategies. The inputs are generated to test all possible operating modes and edge cases.

# <span id="page-51-15"></span><span id="page-51-14"></span><span id="page-51-13"></span><span id="page-51-12"></span><span id="page-51-11"></span><span id="page-51-10"></span><span id="page-51-9"></span><span id="page-51-8"></span><span id="page-51-7"></span><span id="page-51-6"></span><span id="page-51-5"></span><span id="page-51-4"></span><span id="page-51-3"></span><span id="page-51-2"></span><span id="page-51-1"></span><span id="page-51-0"></span>3. Results List of devices (In cooperation with Gr¨omer, [2021\)](#page-87-0) Table44



# Table 3.1.: Lab experiment: Equipment and devices

<span id="page-52-5"></span><span id="page-52-4"></span><span id="page-52-3"></span><span id="page-52-2"></span>List of chemicals and materials (In cooperation with Grömer, [2021\)](#page-87-0)



## Table <sup>3</sup>.2.: Lab experiment: Materials and chemicals

## **Copper sulfate pentahydrate**[1](#page-52-0)



H302: Harmful if swallowed.

<sup>H</sup>318: Causes serious eye damage.

<sup>H</sup>410: Very toxic to aquatic life with long lasting effects.

P273: Avoid release to the environment.

 <sup>P</sup>280: Wear protective gloves/protective clothing/eye protection/face protection.

P305 <sup>+</sup> <sup>P</sup><sup>351</sup> <sup>+</sup> <sup>P</sup>338: IF IN EYES: Rinse cautiously with water for several minutes. Remove contact lenses, if presen<sup>t</sup> andeasy to do. Continue rinsing.

P313: Get medical advice/attention.

## **Sulfuric acid**[2](#page-52-1)

<span id="page-52-1"></span><span id="page-52-0"></span>

<sup>H</sup>290: May be corrosive to metals.

<sup>H</sup>314: Causes severe skin burns and eye damage.

<sup>P</sup>280: Wear protective gloves/protective clothing/eye protection/face protection.

P301 <sup>+</sup> <sup>P</sup><sup>330</sup> <sup>+</sup> <sup>P</sup>331: IF SWALLOWED: rinse mouth. DoNOT induce vomitting.

P303 <sup>+</sup> <sup>P</sup>3<sup>61</sup> <sup>+</sup> <sup>P</sup>353: IF ON SKIN (or hair): Remove/Take off immediately all contaminated clothing. Rinse skin withwater/shower.

P305 <sup>+</sup> <sup>P</sup><sup>351</sup> <sup>+</sup> <sup>P</sup>338: IF IN EYES: Rinse cautiously with water for several minutes. Remove contact lenses, if presen<sup>t</sup>and easy to do. Continue rinsing.

1[http://gestis.itrust.de/nxt/gateway.dll/gestis\\_de/491473.xml](http://gestis.itrust.de/nxt/gateway.dll/gestis_de/491473.xml)2[http://gestis.itrust.de/nxt/gateway.dll/gestis\\_de/001160.xml](http://gestis.itrust.de/nxt/gateway.dll/gestis_de/001160.xml)

45

# 3.2. Electroplating and Predictions

## 3.2.1. Lab-scale experiment

The lab-scale experiment is a cooperation with Grömer, [2021](#page-87-1).

## Lab preparation

The first thing to do in the lab is the setup of the heated water baths (Table [3](#page-51-3).1.1, Table 3.1.2, Table 3.1.3, Table 3.1.4). They need up to 20 minutes to reach and stabilize the set temperature. The second step is to clean the utilities such as beaker glasses (Table 3.1.[19](#page-51-4)), stirrer (Table 3.1.[16](#page-51-5)), and tweezers (Table 3.1.[20](#page-51-6)).

## Production of the electrolyte

According to Jelinek, [2013](#page-87-2), the electrolyte should have a mass concentration of  $100\,\mathrm{g\,L^{-1}} < \beta(\mathrm{CuSO}_4) < 200\,\mathrm{g\,L^{-1}}.$  A batch of electrolyte has 760 mL. This is enough liquid to do four electroplatings parallely. First,  $m(CuSO_4 \cdot 5H_2O) = 133.5g$  $m(CuSO_4 \cdot 5H_2O) = 133.5g$  $m(CuSO_4 \cdot 5H_2O) = 133.5g$  copper sulfate pentahydrate (Table 3.2.1) are weighed. In a second step the CuSO<sub>4</sub>  $\cdot$  5 H<sub>2</sub>O is dissolved in  $m(H_2O) = 750 g$ destilled water (Table 3.[2](#page-52-3).4). The solution homogenizes for 10 minutes using a magnetic stirrer (Table 3.1.[11](#page-51-7)). By adding  $V(H_2SO_4) = 10$  mL sulfuric acid (Table [3](#page-52-4).2.3) the pH-value is set to  $pH = 4$ . After a further homogenization the electrolyte was divided into beakers and tempered in the heated water baths.

## Prepare the copper plates

The copper plates need a few cleaning steps each time before electroplating. In the first step, the copper plates are cleaned with a sponge combined with a scouring pad. It is a mechanical cleaning step to remove copper patina and anode sludge. The copper patina is the product of the natural aging of

copper, depending on weather conditions. It is a copper-(carbonate-sulfatechloride)-hydroxide mixture. Anode sludge contains all non-dissolvable contaminations of the copper plate. The mechanical cleaning with the sponge takes about five minutes per copper plate. The best indicator is when the whole plate has a light copper color (pale pink color). Then the copper plates are rinsed with water. The second cleaning step is ultrasonic cleaning in a ultrasonic bath (Table 3.1.[14](#page-51-8)). Perfume-free soap (Table [3](#page-52-5).2.5) is added to the bathwater as a tenside to remove grease residues. Then the copper plates are rinsed with water again. The last step is the acid cleaning step. Sulfuric acid (Table [3](#page-52-4).2.3) cleans the copper plates' surfaces and dissolves surface contaminations. Calculating the current density (Equation [2](#page-25-0).2) needs the coated copper plate area. Insulation tape swathed the copper plate to have a defined area. The area is then measured with a caliper (Table 3.1.[18](#page-51-9)). The last step is the labeling and the weighting of the copper plates.

### Electroplating

After preparation, labeling, and documentation, the electroplating can start. Crocodile clips (Table 3.1.[21](#page-51-10)) connect the copper plate with the cables. The lab power supplies (Table [3](#page-51-11).1.5, Table [3](#page-51-12).1.6, Table [3](#page-51-13).1.7, Table [3](#page-51-14).1.8) deliver the energy for the electroplating process. The tempered beaker glasses hold the copper plates separated from each other. Thus they do not touch each other and produce a short circuit. The electrolyte must cover the marked area where the material should deposit. Figure 3.[11](#page-55-0) shows the schematic representation of the setup. The lab power supplies produce the energy for the reaction. The parameters are set according to the experiments' plan. Each experiment has its own duration time, current density, and temperature. During the electroplating process, measurements take place. For description, see the Documentation section below. After the duration time, the process stops by turning off the lab power supply. The operator removes the copper plates from the beaker glass, rinses them with water, and dries them with a hairdryer (Table 3.1.[13](#page-57-0)). Figure 3.[12](#page-56-0) and Figure 3.13 show pictures taken in the lab.

<span id="page-55-0"></span>





The upper part shows a view from above. It shows a heated water bath containing the beaker glass where the electroplating takes place. It also shows the arrangement of the copper plates and the wiring. The second figure shows a frontal view of the electroplating process. On the left-hand side, it shows the beaker glass with the copper plates and the wiring. It also shows the isolation tape in green, which separates the processed copper area from the rest of the plate. The right-hand side shows a state-of-the-art lab power supply and the connections to the copper plates.

# 3.2. Electroplating and Predictions

<span id="page-56-0"></span>

Figure 3.12.: Experimental setup The picture shows the setup from Figure 3.[11](#page-55-0) in the lab.

<span id="page-57-0"></span>

Figure 3.13.: Electroplating result

The picture shows the result of the electroplating. In the top mid, it shows the imprint of the holder. At the edges, non-perfectly linear field lines produce a higher deposition than on the surfaces. On the surface is plane and equally distributed deposition.

<span id="page-58-0"></span>

## Documentation

Figure 3.14.: Documentation template

Figure 3.[14](#page-58-0) shows the template for the documentation. It contains all relevant information for the evaluation. The experiment number (VersuchsNr.) is a sequential number to identify and separate the individual experiments. In the target (SOLL) row, the parameters for the experiment are entered. It holds the duration time (min.), the temperature (°*C*), and the current density( $A/dm$ <sup> $\wedge$ </sup>2). The current density depends on the area and the current (see Equation [2](#page-25-0).2) and cannot be set directly. Thus the current must be calculated. The table in the middle of the template holds relevant information about the copper plates. It holds the label (Nummer) of the plates, their start (Masse Start  $[g]$ ), and end (Masse Ende  $[g]$ ) mass, and the area (Fläche Kathode  $\lceil cm^2 \rceil$ ) which is in contact with the electrolyte. The copper plate area's length and width are used to calculate the area which is in contact with the electrolyte. The last six lines hold information for the electroplating process. It holds the electrolyte's temperature at the beginning (Temp 1[°*C*]), in the middle (Temp 2[°*C*]), and at the end (Temp 3[°*C*]) of the process. The same holds for the amperage  $(I_1[A], I_2[A], I_3[A]$ ). Further, the template has a field for the voltage (U [*V*]), the number of the lab power supply (Netzteil), the start (Startzeit) and end (Endzeit) time, and the duration (Dauer [*min*−]) as a result of them. The results of the experiment are in Appendix  $\overline{C}$  $\overline{C}$  $\overline{C}$  (original values) and in Appendix  $\overline{B}$  $\overline{B}$  $\overline{B}$  (processed values).

## Evaluation

The evaluation uses different methods to calculate the layer thickness and the energy consumption of the electroplating process. The original data (see Appendix  $\overline{C}$ ) are transferred to a spreadsheet. The transfer is validated by reference calculations, which are in a certain range or calculated from the data. Per definition of the experiment criteria the current density must be in range of 0.5 A dm<sup>-2</sup> to 5 A dm<sup>-2</sup> (Unruh, [2016](#page-89-0)). The plating area is also a fixed value and is between  $25 \text{ cm}^2$  and  $35 \text{ cm}^2$ . The last criterion is the mass difference. Optimally it is zero, but caused by measuring errors, this does not always hold. So an epsilon  $\varepsilon = 0.07$  g is added to compensate for the errors.

Figure 3.[14](#page-61-0) shows the distributions of the influencing parameters in seperated charts. Each parameter has its own bar chart. Figure  $3.15(a)$  $3.15(a)$  shows the distribution of the current density. The value should be between  $0.5 \text{ A dm}^{-2}$ and  $5 \text{A} \text{dm}^{-2}$ . 18 results ( $\sim 9\%$ ) are out of range. They are still used in the evaluation because they are valid results. The boundaries were set in the lab experiment preparations to get a rough idea of the values. The boundaries for the duration are 10 min as the minimum and 90 min as the maximum. Figure 3.15[\(b\)](#page-60-1) shows the distribution of the duration. The data set has 4 out layers (∼ 2%). The results are valid and used in the evaluation. The temperature distribution does not have any out layers. Figure  $3.15(c)$  $3.15(c)$  shows its distribution. During the experiments, only three results are invalid. One was a documentation error, and two results are invalid because of technical issues.

The valid data sets are the basis for further calculations and evaluations. The first approach is to use Equation [2](#page-24-0).1 to calculate the layer thickness from the parameters and compare them to the gathered data. The coefficient of determination *R* 2 (Rinne and Ickler, [1986](#page-88-0)) determines the comparison.

<span id="page-60-0"></span>

<span id="page-60-1"></span>(b) Duration distribution

<span id="page-61-1"></span><span id="page-61-0"></span>

Figure 3.14.: Distribution of experiment parameters

The lab experiments consist of over 200 experiments. In each experiment, the duration, the current density, and the temperature are chosen randomly within given boundaries. The bar charts show the distribution of the parameters.

$$
R^{2} = \frac{\sum (\hat{y}_{i} - \bar{y})^{2}}{\sum (y_{i} - \bar{y})^{2}}
$$
 (3.16)

where:

 $\hat{y}_i$  = Predicted measurement  $y_i$  = Single measurement  $\bar{y}$  = Mean of *y* 

<span id="page-61-2"></span>A further step uses neural networks and other machine learning methods to predict the layer thickness. Chapter [2](#page-25-1).4 gives an overview of the used methods. The methods take the duration, temperature, and current density as input. All methods use supervised learning, so the layer thickness (resp. the mass of deposit copper) is also a parameter for the training. The trained models are used to predict further measurements. The results of the models can be found in Chapter [3](#page-62-0).2.2. The methods described map the process well. The last step is the calculation of the needed electric energy for the procedure. The electric power *P*, and the electric potential energy *W* (Plamann and Schulz, [2016](#page-88-1)) can be calculated from the existing data.

$$
W = P * t \tag{3.17}
$$

where:

*W* = Electric potential energy  $P =$  Electric power, see Equation 3.[18](#page-62-1)  $t =$ Duration time

$$
P = U * I \tag{3.18}
$$

<span id="page-62-1"></span>where:

*P* = Electric power  $U =$  Voltage  $I =$ Electric current

The measurements and the calculated data are used to train a model, as well. This time the predicted value is the electric potential energy. The results of these models are in Chapter [3](#page-66-0).2.3.

## <span id="page-62-0"></span>3.2.2. Coating thickness prediction

Appendix [B](#page-102-0) shows 2 tables. The first table contains the documented data from the lab experiment. The second table contains the calculated values from the first table. The first evaluation of the data is the calculation of the coating thickness in two ways. The first approach is the expected coating thickness *t<sup>e</sup>* , using Equation [2](#page-24-0).1. The second approach is to calculate the real coating thickness  $t_r$  from the deposit mass of copper and the coating area.

$$
t_r = \frac{\Delta m_C}{\varrho * A} \tag{3.19}
$$

where:

 $t_r$  = Real coating thickness  $\Delta m_C$  = Mass of deposit copper

 $\varrho =$  Density of copper

 $A =$ Coating area

In more than 90 % of the experiments, the deviation between the real and the expected coating thickness is smaller than 15 %. The *R* <sup>2</sup> over the whole data set is 0.98, which is very close to 1 and, therefore, excellent. The calculated real coating thickness *t<sup>r</sup>* is the label for the documented data. With data and the corresponding label, models are trained to predict the coating thickness. The data set is divided and 70% of the data are used as training data. The used models are described in Chapter [2](#page-25-1).4.

<span id="page-63-0"></span>

Figure 3.15.: Coating thickness model A machine learning model needs the duration, the current density, and the temperature to predict the coating thickness.

Figure 3.[15](#page-63-0) shows the input features and the output/label for the created models. Chapter [4](#page-77-0).2.1 holds the interpretation of the coating thickness prediction results. To get comparable and reproducible results, the random state is fixed for all methods which use it.

## Linear regression

The linear regression does not need any further information than the data as input. The *R* <sup>2</sup> value determined the similarity between the predicted and the real values. It is applied to the test set and the training set. Table [3](#page-64-0).3 shows the results on the test set and the training set.

## 3.2. Electroplating and Predictions

Table 3.3.: Coating thickness prediction: Linear regression model			
		$R^2$ on test set $\left  R^2$ on training set $\right $	
Linear regression $\vert$	0.812	0.777	

<span id="page-64-0"></span>Table 3.3.: Coating thickness prediction: Linear regression model

### Polynomial regression

The polynomial regression needs the maximal degree of the polynomial as an additional parameter. The model is created seven times with different maximal degree parameters. Setting the degree to one, the polynomial regression model becomes a linear regression model. Table [3](#page-64-1).4 shows the results on the test set and the training set. With a degree of  $7$  the  $R^2$  becomes negative. This happens when the model is worse than the null hypothesis. This means that the model fits the data worse than a straight line.

<span id="page-64-1"></span>Table 3.4.: Coating thickness prediction: Polynomial regression models

Maximal degree	$\overline{R}^2$ on test set	$\overline{R}^2$ on training set
	0.812	0.777
2	0.982	0.974
	0.977	0.979
	0.975	0.983
	0.974	0.986
	0.898	0.989
	$-1.357$	0.993

## Decision tree regression

The Decision tree regression needs the maximal depth of the tree as an additional parameter. The model is created 19 times with different maximal depth parameters. Table [3](#page-65-0).5 shows the results on the test set and the training set.



<span id="page-65-0"></span>Table 3.5.: Coating thickness prediction: Decision tree regression models

## Random forest regression

The random forest regression does not need any further information than the data as input. Table [3](#page-65-1).6 shows the results on the test set and the training set.

<span id="page-65-1"></span>Table 3.6.: Coating thickness prediction: Random forest regression model

		$\sqrt{R^2}$ on test set $\sqrt{R^2}$ on training set $\sqrt{R^2}$
Random forest regression	0.959	0.994

## Neural networks

<span id="page-66-1"></span>The neural network consists of three hidden layers with 30 neurons each. The input layer is also a 30 neurons layer. Table [3](#page-66-1).7 shows the results on the test set and the training set.



## <span id="page-66-0"></span>3.2.3. Electric potential energy prediction

To calculate the real electric potential energy Equation 3.[17](#page-61-2) and Equation 3.[18](#page-62-1) are used. The results are the labels for the data models. Further, the data models get the duration, the temperature, the amperage, and the coating area as inputs. Each model used 70% of the data as training data. Figure 3.[16](#page-66-2) shows the structure of the data model. Chapter [4](#page-78-0).2.2 holds the interpretation of the power prediction results. To get comparable and reproducible results, the random state is fixed for all methods which use it.

<span id="page-66-2"></span>

Figure 3.16.: Electric potential energy model A machine learning model needs the duration, the amperage, the voltage, the coating area, and the temperature to predict the electric potential energy.

## Linear regression and Polynomial regression

As stated in the last Chapter, linear regression is the same as a polynomial regression with one as its maximal degree. Thus both evaluations are combined in one section. The maximal degree parameter is varied from one to eight. Table [3](#page-67-0).8 shows the results for varying degrees.

Maximal degree	$R^2$ on test set	$R^2$ on training set
$1 =$ Linear regression	0.816	0.813
2	0.973	0.988
	0.999	1.000
	1.000	1.000
	1.000	1.000
	0.997	1.000
	0.959	1.000
	$-0.694$	1.000

<span id="page-67-0"></span>Table 3.8.: Electric potential energy: Linear regression and polynomial regression models

## Decission tree regression

The Decision tree regression needs the maximal depth of the tree as an additional parameter. The model is created 19 times with different maximal depth parameters. Table [3](#page-68-0).9 shows the results on the test set and the training set.

## 3.2. Electroplating and Predictions

<span id="page-68-0"></span>

Table 3.9.: Electric potential energy: Decision tree regression

## Random forest regression

The random forest regression does not need any further information than the data as input. Table 3.[10](#page-68-1) shows the results on the test set and the training set.

Table 3.10.: Electric potential energy: Random forest regression model

<span id="page-68-1"></span>

		$\mathbb{R}^2$ on test set $\mathbb{R}^2$ on training set
Random forest regression	0.887	0.987

### Neural network

<span id="page-69-0"></span>The neural network consists of three hidden layers with 30 neurons each. The input layer is also a 30 neurons layer. Table 3.[11](#page-69-0) shows the results on the test set and the training set.



# 3.3. Function-based approximation of data

This section presents the structure of the program and its results. The approximation is written in Python. The script has the following steps:

- Load data
- Calculate function parameters
- Calculate off-times
- Predict data for the future

## Load data

In the course of this thesis, a program was developed to perform the function-based approximation automatically. The program has to load two files. The first file is a comma-separated values file (CSV file). It contains the data and the corresponding timestamps. The program uses this file to calculate the parameters for the approximation functions and to draw compare graphs. The second file contains information about the off times of the plant. The program uses this data to decide whether the plant is in production mode or turned off. Further, the data is needed to determine the point when the plant has to start heating, so it reaches production temperature when production begins.

## Calculate function parameters

In the first place, the program calculates the default value for production. This value is the set point of the plant. The only known fact at this point is that the plant is mostly in production mode and only in about 15% (Weekend and Maintenance) off. So the program calculates the median *M* and the standard derivation *SD* of the data from the CSV-file. The median method sorts all *n* values from lowest to highest and picks the value located at index *n*/2 as its result. It is very likely to choose a value that belongs to the production phase because the production phase is about 85% of the time. To be sure that the value is a production phase value, the standard deviation is calculated, and then the program filters the whole data set and removes every value which is not in the range  $M \pm SD$ . On the filtered data set, the median is again calculated. This double-check removes every value which does not belong to the production phase and results in a value used as set point *SP* and the corresponding *SD*. In a second step, the variation around the *SP* is modeled. The deviation around the *SP* is shaped like a sine function. When the value gets too low, the bath starts heating and stops when it reaches a turn-off value. This up and down makes the data look like they are oscillating, and the sine function represents this behavior. The program used a general sine function

$$
f(t) = A * sin(\omega * t + \phi) + \alpha
$$
 (3.20)

where:

 $f(t)$  = Value at a given time *t* 

 $A =$ The amplitude of the sine function

*ω* = Correction factor for *t*

- $\phi$  = The shift of the phase
- *α* = Offset in *y* direction

The program fits the parameters to the given data using Scipys (Virtanen et al., [2020](#page-89-1)) optimization functions. This results in parameters *A*, *ω*, *φ*, and  $\alpha$  to describe the sine function. In the last step, the parameters for the cooldown and the heat up are calculated. This step aims to find parameters for the Equation [2](#page-32-0).6 and Equation [2](#page-32-1).7. It first identifies the sequences where

the cooldown and the heat up happen by filtering the data and removing the data at the  $SP \pm SD$  and then finds the most prolonged phase where it happens. This way it gets the most precise values for the parameter calculation. Then the program again fits the equations to the data using the Scipy package. By doing this, the parameters for cooldown and heat up are determined.

## Calculate off-times

The off-times are mostly determined by the maintenance and the weekends. During the creation of the offtime-object, the program calculates already the point when heat up has to start. For this it solves Equation [2](#page-32-0).6 and Equation [2](#page-32-1).7 for the time when they meet at the same point. It results in

$$
t = \frac{\mathcal{W}\left(\frac{e^{c_{c}*(T_{X}-T_{A})}}{c_{h}} * c_{c}*(T_{0}-T_{A})\right) * c_{h} - c_{c}*(T_{X}-T_{A})}{c_{c} * c_{h}}
$$
(3.21)

where:

- $t =$  Time where the two functions are equal
- $T_0$  = The temperature of the fluid at the beginning of cooldown
- $T_A$  = The temperature of the ambient
- $T_X$  = Fictive temperature where heat up has to start at begin of cooldown A coefficient affected by the volume of the fluid, the surface,
- $c_c =$ thermal capacity, and the density of the fluid
- $c_h$  = A coefficient affected by the volume of the fluid, the surface,  $c_h$  = thermal canacity and the density of the fluid
- thermal capacity, and the density of the fluid
- $W =$  Lambert W function

The precalculation makes the prediction less compute intensely. Instead of checking each time in which phase of the off-time a given timestamp is, the program has to check if the timestamp is smaller than *t*, then it is in the cooldown phase; otherwise, it is in the heat-up phase.
#### Predict data for the future

After all the preparations, the prediction is an easy task. For each given timestamp, the program checks if it is in an off-time or not. If it is not, then it uses the calculated sine parameters to get the prediction value. If it is an off-time phase, then the program checks if it is before or after the turning point and selects the cooldown curve with the calculated cooldown parameters if it is before the turning point. IF not it uses the heat-up function and the computed parameters. Figure 3.[17](#page-72-0), Figure 3.[18](#page-73-0) and Figure 3.[19](#page-74-0) show the results of the prediction in general and in detail.

<span id="page-72-0"></span>

#### Timestamp [ms]

Figure 3.17.: Real data vs. prediction data

The orange curve shows the real data, and the blue curve shows the predicted data. The predicted data follows the real data quite well. There is a gap in the real data from a recording error in the middle of the graph, but the prediction does still work because it does not depend on the real data. For more details, see Figure 3.[18](#page-73-0) and Figure 3.[19](#page-74-0).

<span id="page-73-0"></span>

Figure 3.18.: Real data vs. prediction data: Off-time phase

The orange curve shows the real data, and the blue curve shows the predicted data. This graph is a more detailed view of the off-time phase. The prediction is pretty accurate. Only in the heat-up phase, there is a small drop in the real data, which cannot be considered in the forecast.

<span id="page-74-0"></span>



The orange curve shows the real data, and the blue curve shows the predicted data. This graph is a more detailed view of the production phase. It shows the real data and the approximated sine function. The approximations fit the real data well. In some cases, the prediction is shifted, and the real data and the prediction data are inversed. It is not an actual fault because it can even happen in actual production that the data shift, caused by ambient conditions or shortstops in production.

### 4. Discussion

The key feature of the DET model is the imitation of the real plant. It is the main feature for the digital model's energy calculation and makes the twin more comparable with its real antagonist. The thesis's main parts are the DT model, the data models for the electroplating as preparatory study for the DT model, and sensor data approximation. All aspects brought good results. The DT model handles various control mechanisms that are essential for the operation of the plant and the correctness of the model. The data model for the electroplating works well. They show accurate results and fit the actual data well. Since the plant uses quite the same plating method, a transfer from the lab-scale model to the real-world model is likely and can be done in follow-up work. The current solution for the sensor data approximations works well for liquid tanks, which have cool-down and heat-up phases. Adopting other scenarios is possible, but it is an effort to find proper functions and optimize them to fit the data. The DET project is ongoing; thus the methods and tools developed in this thesis are a vital part of the project.

### 4.1. Digital Energy Twin Model

The development of the DET model is an ongoing process, and many parts of the model are still in development or haven't even started. The primary air consumers, the drilling machines, were not modeled yet. Due to these circumstances, it is impossible to test the control models developed in the thesis with real data inputs; thus, it is impossible to show the control logic under real conditions. The results show the control module structure and experiments with test data to see if the implementation works as expected. Also, the water consumers and water return devices are not implemented in

#### <span id="page-77-1"></span>4. Discussion

the DET model by now. This affects the tank logic and the overflow logic. All control signals generated by these two modules depend on the water level in the tanks. So the evaluation of the model also took place with test data instead of real data. When the other parts will be finished, another test with real data should evaluate the actual scenario's control models' functionality. With new insights and further developement of the other parts it can happen that the created models need to be adapted, but the underlaying concept stays the same.

### 4.2. Electroplating and Predictions

Overall, the data models work well (see Chapter [4](#page-77-0).2.1 and Chapter 4.[2](#page-78-0).2), but there are also limitations by the mathematical foundation of the models. The training set heavily limits the decision tree regression and the random forest regression. Both regressions cannot predict outside of the limits of the training set parameters. For them, it is not possible to see patterns in the data. A pattern would be a direct proportion between the duration time and the coating thickness in the electroplating data set. The longer the coating takes, the thicker the layer becomes. Also, polynomial regression is not aware of such outlying data. The ML cannot consider physical constraints, which are not covered by the training set. A perfect example would be the whole consumption of the anode. If there is no anode left, the reaction cannot occur anymore, and then the duration time can be infinite long without depositing any copper. But it is not possible to take all these effects into account. The data models represent a working, non-malicious environment. In future work, it would be interesting to integrate physical knowledge in ML algorithms. Rueden et al., [2019](#page-88-0) show approaches for this.

### <span id="page-77-0"></span>4.2.1. Coating thickness prediction

The *R* <sup>2</sup> value is a relevant measurement and a good indicator of the quality of data models. Chapter [3](#page-62-0).2.2 shows this information for the used data models. Table [4](#page-78-1).1 shows the best results for each of the data models on the test set.

#### 4.2. Electroplating and Predictions

<span id="page-78-1"></span>

Data model		$R^2$ on test set   $R^2$ on training set
Polynomial regression (degree $= 2$ )	0.982	0.974
Decision tree regression (depth $= 7$ )	0.940	0.997
Random forest regression	0.949	0.994
Neural networks	0.946	0.991

Table 4.1.: Best results for the data models

Linear regression and polynomial regression are combined into one line in the table because the linear regression is a polynomial regression of the degree at most one. Polynomial regression performs well for this data model, and even a small polynomial degree has excellent results on the test set and the training set. The other methods also perform well on the data and are very close to the polynomial regression results. Decision tree regression and random forest regression are very similar because they are based on the same background. Separating the data into a test set and a training set is crucial to detect and reduce overfitting. Most data models overfit when they get too confident on the training data. This especially happens to higher degrees for the polynomial regression.

### <span id="page-78-0"></span>4.2.2. Electric potential energy prediction

The *R* <sup>2</sup> value is a relevant measurement and a good indicator of the quality of the method. Chapter [3](#page-66-0).2.3 shows this information for the used data models. Table [4](#page-78-2).2 shows the best results for each of the data models on the test set.

<span id="page-78-2"></span>

Data model		$R^2$ on test set   $R^2$ on training set
Polynomial regression (degree $=$ 4)	1.000	1.000
Decision tree regression (depth = $9$ )	0.895	1.000
Random forest regression	0.887	0.987
Neural networks	0.939	0.990

Table 4.2.: Best results for the data models

#### <span id="page-79-0"></span>4. Discussion

For the electric potential energy, a polynomial regression has the best performance as well. The model is based on more parameters than the coating thickness model, so it is not surprising that also the polynomial regression needs a higher degree to have the best performance. The other data models also perform excellently on the training set, but they have weaknesses in finding the real context of the features, and so they do not perform so well on the test data set. Decision tree regression and Random forest regression have a lousier performance than neural networks and polynomial regression. Neural networks work the same for both data model scenes.

### 4.3. Function-based approximation of data

The method approximated the data well. The main point of good approximation is to know the off-times of the plant. For past data, it is a look at the data to see when production ends and when it starts again. Nevertheless, for future predictions, this is not possible since the data does not exist yet. A reference point is the plant's shift plan, which is planned for the future and gives information on when production is on and when not. It leads to a rough estimation that works, but it negatively influences future predictions based on the approximation; they get inaccurate. The further the prediction is in the future, the more inaccurate it gets. A compromise is a limitation of future predictions. For example, limit them to two weeks. Thus the prognosis does not get too imprecise.

#### *A fool with a tool is still a fool!* – Parker and HP OpenView Business Unit, [2001](#page-88-1)

Function-based approximation of data is an excellent method to generate predictions. Like every other prediction method, it has some advantages and some drawbacks. The structure is apparent, and there are no hidden parts that are defined by a training algorithm. The underlying data determine every parameter, and the connections are visible. It is easy to see how a program calculates the result of the prediction. It makes function-based approaches easy to understand and, therefore, suitable, controllable and adaptable. But this also leads to drawbacks. To create such a function-based

approximation, one needs to understand the system and the underlying data. This is an easier task on data with a physical background than on natural-based data like speech recognition. On natural-based data, functionbased approximations come to their limits. In general, the user has to know how to use the tools and where the boundaries are.

## 5. Conclusion

This thesis contributed to the development of Digital Energy Twins for Industry 4.0. DETs are important to optimize systems for various parameters without affecting the real production system. This thesis addressed three research topics. The research aimed to calculate the electroplating process's energy consumption with ML methods, to optimize sensor data usage by function-based approximation, and further developed the DT model with Modelica.

Based on the lab experiment and the collected data, the energy consumption was calculated. With various ML techniques, it is possible to predict the energy consumption for an electroplating process. The ML methods' main features are the voltages, amperage, temperature, coating area, and the duration of the plating process. A data model fed with these features creates a prediction very close to the real energy consumption. The *R* <sup>2</sup> value of the polynomial regression with a degree of two is 1.000. It is an excellent value, and  $R^2$  is the right measurement for the method's quality. The chosen method provides a good estimate of the energy consumption. The accuracy is sufficient for the whole electroplating plant's energy calculation since all components scatter around their real values.

The function-based approximation of sensor data is a crucial method for the development of digital energy twins. It has some main advantages. First of all, it allows creating an interface for the DT model and gather the data in a computationally effective way. Further, the approximation also allows predicting data in the future. It only needs information about the weekends' shutdown or machine maintenance. It is possible to generate this information with the shift plan, the maintenance plan, and the plant's utilization schedule. Since these plans only contain rough estimations when the events happen, and unplanned events entail maintenance, predictions in the far future are inaccurate. The created program for function-based approximation can make these predictions, but it is not recommended

#### 5. Conclusion

to predict too far into the future. Up to two weeks should be reasonable because it normally contains four events (two times weekends and two times maintenance), and then the expected behavior does not differ too much from the real one. Thus further calculations on the predictions, like energy calculations for the tempering, are possible without the risk of heavy miscalculation.

The DT model is a work in progress. The research in this thesis is a integral part of the whole model, impacting the DT model's work. The control models allow a precise control of the individual machines. Since each device has its own control model, it is easy to maintain and test. The delays in the control strategies cover the start-up time for the real machines. The start-up and run-on time of the the real machine are the most significant differences between the model and the real-world object, because Modelica libraries do not take such effects into account.

Here are some recommendations for the project and the handling of the tool created within this thesis.

The basis for the function-based approximations tool is the sensor data. The collection of the data started in October 2020. The tool needs the data to calculate the parameters for the functions. Even if there is already enough data for the tool to work, it is still recommended to collect further data for at least a year or even more. The current measurements were taken in autumn and winter, so the temperature was low, and the humidity was high. These or other seasonal effects may affect the measurements and lead to fluctuations in the data. It could then affect the tool's calculations because real data changes, but the function-based approximation parameters stay the same.

The prediction of the function-based approximations tool is based on production plans. The plans cannot take unexpected failures into account. The further the prediction reaches into the future, the higher the chance that an unexpected failure happens. It leads to the recommendation that the prediction should not be longer than two weeks into the future. This recommendation is not based on expert assessment within the project.

The energy models from the lab experiment are a standalone tool. The validation of this tool was done with the lab data. When the model was developed, the plant's data was not available, so it was impossible to crossvalidate. As described in Chapter [3](#page-53-0).2 the electroplating process is similar, but the lab process still differs from the plant process. Before using the electroplating data models created, it is recommended to validate the model's results with the plant's energy consumption.

It is still a long way for the project to cross the finish line. The presented thesis supports the success of the project in various ways. The modeling part allows the model to generate more accurate results. With more complex control sequences, the model gets closer to the real system and allows a better and more precise energy calculation. The data models for the electroplating process are an alternative for the real process in the plant. They allow the simplification of a complex process without losing accuracy. Since not the process but the process's energy consumption is the project's primary goal, it is a suitable simplification that saves a lot of modeling work to represent the process. The function-based approximation predicts data for the future. Having information for the future also allows calculating future energy consumptions and simulating future events with the DET model, and this is one of the main goals of a DET. All this drives the project towards its goal of creating a Digital Energy Twin.

## **Bibliography**

- Aggarwal, Charu C. (2020). *Linear Algebra and Optimization for Machine Learning*. 1st ed. 20. Springer (cit. on p. [18](#page-25-0)).
- Bendel, Oliver (2019). *Industrie* 4.0. URL: https://wirtschaftslexikon. [gabler . de / definition / industrie - 40- 54032 / version - 368841](https://wirtschaftslexikon.gabler.de/definition/industrie-40-54032/version-368841) (Accessed on  $10/22/2020$  $10/22/2020$  $10/22/2020$  (cit. on p. 2).
- Black, Newton Henry and Harvey Nathaniel Davis (1922). *Practical Physics: Fundamental Principles and Applications to Daily Life*. Macmillan (cit. on pp. [14](#page-21-0), [15](#page-22-0)).
- El Zein, Ahmad L and Nour A Chehayeb (2015). "The effect of greenhouse gases on earth's temperature." In: *International Journal of Environmental Monitoring and Analysis* 3 (cit. on p. [1](#page-8-0)).
- ENERTEC Naftz & Partner GmbH & Co KG (2020). *Ubersicht Werk Hinterberg ¨* (cit. on p. [27](#page-34-0)).
- Flauger, Jürgen et al. (2020). *Steigender Energiebedarf: Deutschland droht die Okostrom-Lücke.* URL: [https://www.handelsblatt.com/unternehmen/](https://www.handelsblatt.com/unternehmen/energie/energie-steigender-energiebedarf-deutschland-droht-die-oekostrom-luecke/25385468.html) energie/energie-steigender-energiebedarf-deutschland-droht[die-oekostrom-luecke/25385468.html](https://www.handelsblatt.com/unternehmen/energie/energie-steigender-energiebedarf-deutschland-droht-die-oekostrom-luecke/25385468.html) (Accessed on 10/24/2020) (cit. on p. [2](#page-9-0)).
- Fritzson, P and B Thiele (2016). "Introduction to Object-Oriented Modeling, Simulation, Debugging and Dynamic Optimization with Modelica using OpenModelica." In: *Simulation* (cit. on p. [10](#page-17-0)).
- Fritzson, Peter (2011). *Introduction to modeling and simulation of technical and physical systems with Modelica*. John Wiley & Sons (cit. on p. [10](#page-17-0)).
- Gaul, Bernhard (2018). *Osterreichs Klimaplan f ¨ ur¨ 2030: Milliardenstrafe aus Brüssel droht.* URL: [https://kurier.at/politik/inland/oesterreichs](https://kurier.at/politik/inland/oesterreichs-klimaplan-fuer-2030-milliardenstrafe-aus-bruessel-droht/400346833)[klimaplan - fuer - 2030 - milliardenstrafe - aus - bruessel - droht /](https://kurier.at/politik/inland/oesterreichs-klimaplan-fuer-2030-milliardenstrafe-aus-bruessel-droht/400346833) [400346833](https://kurier.at/politik/inland/oesterreichs-klimaplan-fuer-2030-milliardenstrafe-aus-bruessel-droht/400346833) (Accessed on 11/02/2020) (cit. on p. [5](#page-12-0)).

#### <span id="page-87-0"></span>**Bibliography**

- Glaessgen, Edward and David Stargel (2012). "The digital twin paradigm for future NASA and US Air Force vehicles." In: *53rd Structures, Structural Dynamics, and Materials Conference*, p. 1818 (cit. on p. [5](#page-12-0)).
- Grieves, Michael (2002). "Conceptual ideal for PLM." In: *Presentation for the Product Lifecycle Management (PLM) center, University of Michigan* (cit. on p. [5](#page-12-0)).
- Grömer, Michael (2021). "Development, validation and simplification of component models for DigitalEnergyTwin" (cit. on pp. [44](#page-51-0)–[46](#page-53-1), [95](#page-102-0)).
- Hasegawa, Tomoko et al. (2018). "Risk of increased food insecurity under stringent global climate change mitigation policy." In: *Nature Climate Change* 8, pp. 699–703 (cit. on p. [1](#page-8-0)).
- International Energy Agency (2019). *Stromerzeugung weltweit nach Energieträger im Jahresvergleich* 1998 *und* 2017 (in Terawattstunden). URL: [https:](https://de.statista.com/statistik/daten/studie/190298/umfrage/stromerzeugung-weltweit-nach-energietraegern-seit-1998/) [/ / de . statista . com / statistik / daten / studie / 190298 / umfrage /](https://de.statista.com/statistik/daten/studie/190298/umfrage/stromerzeugung-weltweit-nach-energietraegern-seit-1998/) [stromerzeugung- weltweit- nach- energietraegern- seit- 1998/](https://de.statista.com/statistik/daten/studie/190298/umfrage/stromerzeugung-weltweit-nach-energietraegern-seit-1998/) (Accessed on  $11/26/2020$  $11/26/2020$  $11/26/2020$  (cit. on p. 2).
- IPCC (Intergov. Panel Clim. Change) (2018). *Global Warming of 1.5 C. An IPCC Special Report on the Impacts of Global Warming of 1.5* ◦ *C Above Pre-Industrial Levels and Related Global Greenhouse Gas Emission Pathways, in the Context of Strengthening the Global Response to the Threat of Climate Change* (cit. on p. [1](#page-8-0)).
- Jelinek, T. W. (2013). *Praktische Galvanotechnik : ein Lehr- und Handbuch. 96 Tabellen im Text, einem speziellen Tabellen-Anhang und einer Ubersicht ¨ "Chemikalien fur die Galvanotechnik" ¨* . 7., aktual. Leuze Verlag (cit. on pp. [12](#page-19-0), [46](#page-53-1)).
- Kanani, Nasser (2009). *Galvanotechnik : Grundlagen, Verfahren und Praxis einer Schlüsseltechnologie*. Carl Hanser Verlag GmbH Co KG (cit. on p. [12](#page-19-0)).
- Kritzinger, Werner et al. (2018). "Digital Twin in manufacturing: A categorical literature review and classification." In: *IFAC-PapersOnLine* 51.11, pp. 1016–1022 (cit. on pp. [6](#page-13-0), [7](#page-14-0)).
- Lasi, Heiner et al. (2014). "Industry 4.0." In: *Business & information systems engineering* 6.4, pp. 239–242 (cit. on p. [2](#page-9-0)).
- Li, Lorraine (2019). "Classification and Regression Analysis with Decision Trees." In: *Towards Data Science*. URL: https://towardsdatascience. [com / https - medium - com - lorrli - classification - and - regression](https://towardsdatascience.com/https-medium-com-lorrli-classification-and-regression-analysis-with-decision-trees-c43cdbc58054)  [analysis-with-decision-trees-c43cdbc58054](https://towardsdatascience.com/https-medium-com-lorrli-classification-and-regression-analysis-with-decision-trees-c43cdbc58054) (Accessed on 12/30/2020) (cit. on p. [22](#page-29-0)).
- Masson-Delmotte, Valérie et al. (2018). "Global warming of 1.5 C." In: An *IPCC Special Report on the impacts of global warming of 1.5 C* 1, pp. 1–9 (cit. on p. [2](#page-9-0)).
- Modelica Association (2020). *The Modelica Association*. URL: [https://www.](https://www.modelica.org/) [modelica.org/](https://www.modelica.org/) (Accessed on 11/19/2020) (cit. on p. [9](#page-16-0)).
- Nocedal, Jorge (2006). *Numerical optimization*. 2. ed. Springer series in operations research and financial engineering (cit. on p. [21](#page-28-0)).
- Owusu, Phebe Asantewaa and Samuel Asumadu-Sarkodie (2016). "A review of renewable energy sources, sustainability issues and climate change mitigation." In: *Cogent Engineering* 3.1 (cit. on p. [1](#page-8-0)).
- <span id="page-88-1"></span>Parker, Lindsay and HP OpenView Business Unit (2001). "A fool with a tool is still a fool." In: *HP Open View* (cit. on p. [72](#page-79-0)).
- Patel, Meghna Babubhai, Jagruti N Patel, and Upasana M Bhilota (2020). "Latest Technology and Future Trends." In: *Applications of Artificial Neural Networks for Nonlinear Data*. IGI Global, pp. 270–273 (cit. on p. [18](#page-25-0)).
- Pedregosa, Fabian et al. (2011). "Scikit-learn: Machine learning in Python." In: *The Journal of machine Learning research* 12, pp. 2825–2830 (cit. on p. [20](#page-27-0)).
- Plamann, Wilfried and Detlef Schulz (2016). *Handbuch Elektrotechnik: Grundlagenund Anwendungen für Elektrotechniker*. Springer (cit. on p. [54](#page-61-0)).
- Ramírez, Hermes et al. (2017). "Modelica Language Specification." In: Data *in Brief* 25 (cit. on p. [9](#page-16-0)).
- Rashid, Tariq (2017). *Neuronale Netze selbst programmieren: Ein verstandlicher ¨ Einstieg mit Python*. O'Reilly (cit. on p. [20](#page-27-0)).
- Reay, Dave et al. (2007). "Climate change 2007: spring-time for sinks." In: *Nature* 446, pp. 727–8 (cit. on pp. [1](#page-8-0), [2](#page-9-0)).
- Rinne, Horst and Günter Ickler (1986). *Grundstudium Statistik*. 2. Aufl., Verlag für Wirtschaftsskripten (cit. on p.  $52$ ).
- Rohith, Gandhi (2018). "Introduction to Machine Learning Algorithms: Linear Regression." In: *Towards Data Science*. url: [https://towardsdatascie](https://towardsdatascience.com/introduction-to-machine-learning-algorithms-linear-regression-14c4e325882a)nce. [com/introduction-to-machine-learning-algorithms-linear-regress](https://towardsdatascience.com/introduction-to-machine-learning-algorithms-linear-regression-14c4e325882a)ion-[14c4e325882a](https://towardsdatascience.com/introduction-to-machine-learning-algorithms-linear-regression-14c4e325882a) (Accessed on 12/28/2020) (cit. on p. [20](#page-27-0)).
- Rojas, Raul (1996). *Theorie der neuronalen Netze: Eine systematische Einführung.* 1. Aufl. Springer (cit. on p. [20](#page-27-0)).
- <span id="page-88-0"></span>Rueden, Laura von et al. (2019). "Informed Machine Learning–A Taxonomy and Survey of Integrating Knowledge into Learning Systems." In: *arXiv preprint arXiv:1903.12394* (cit. on p. [70](#page-77-1)).

#### **Bibliography**

- Sai Ajay, Muktha (2020). "Introduction to Artificial Neural Networks." In: *Towards Data Science*. URL: https : //towardsdatascience.com/ [introduction- to- artificial- neural- networks- ac338f4154e5](https://towardsdatascience.com/introduction-to-artificial-neural-networks-ac338f4154e5) (Accessed on  $01/08/2021$  $01/08/2021$  $01/08/2021$  (cit. on p. 20).
- Tao, Fei et al. (2018). "Digital twin-driven product design, manufacturing and service with big data." In: *The International Journal of Advanced Manufacturing Technology* 94.9, pp. 3563–3576 (cit. on p. [5](#page-12-0)).
- Uhlemann, Thomas H-J, Christian Lehmann, and Rolf Steinhilper (2017). "The digital twin: Realizing the cyber-physical production system for industry 4.0." In: *Procedia Cirp* 61, pp. 335–340 (cit. on p. [3](#page-10-0)).
- Unruh, Jürgen (2016). *Lehrbuch der Galvanotechnik*. Leuze Verlag (cit. on pp. [12](#page-19-0), [13](#page-20-0), [15](#page-22-0), [17](#page-24-0), [52](#page-59-0)).
- Virtanen, Pauli et al. (2020). "SciPy 1.0: fundamental algorithms for scientific computing in Python." In: *Nature methods* 17.3, pp. 261–272 (cit. on p. [63](#page-70-0)).
- Watts, C et al. (2016). "A brewing storm: the climate change risks to coffee." In: *A brewing storm: the climate change risks to coffee.* (cit. on p. [1](#page-8-0)).
- Wetter, Michael et al. (2014). "Modelica buildings library." In: *Journal of Building Performance Simulation* 7.4, pp. 253–270 (cit. on p. [9](#page-16-0)).
- WKO Austria (2020). *Bruttoinlandsprodukt (BIP)*. URL: [https://www.wko.at/](https://www.wko.at/service/zahlen-daten-fakten/BIP.html) service/zahlen-daten-fakten/BIP.html (Accessed on 10/24/2020) (cit. on p. [2](#page-9-0)).
- Wuebbles, Donald J et al. (2017). *Climate science special report: Fourth national climate assessment (NCA4), Volume I* (cit. on p. [1](#page-8-0)).
- Yiu, Tony (2019a). "Understanding Neural Networks." In: *Towards Data Sci*ence. URL: [https://towardsdatascience.com/understanding-neural](https://towardsdatascience.com/understanding-neural-networks-19020b758230)[networks-19020b758230](https://towardsdatascience.com/understanding-neural-networks-19020b758230) (Accessed on 01/08/2021) (cit. on p. [20](#page-27-0)).
- Yiu, Tony (2019b). "Understanding Random Forest." In: *Towards Data Sci*ence. URL: [https://towardsdatascience.com/understanding-random](https://towardsdatascience.com/understanding-random-forest-58381e0602d2)[forest-58381e0602d2](https://towardsdatascience.com/understanding-random-forest-58381e0602d2) (Accessed on 12/30/2020) (cit. on p. [23](#page-30-0)).

# List of Figures





## List of Tables



# List of Reactions



# List of Equations





# Appendix

## Appendix A.

### Overview energy flow

ENERTEC Naftz & Partner GmbH & Co KG created the overview on the next page. It is the basis for the DET model. It shows the structure of the plant and how the energy flows of the machines are connected. The overview has small modifications caused by regulations. The modifications hide the name of the company and the location of the plant. All modifications are marked with three red crosses.



### <span id="page-102-0"></span>Appendix B.

### Lab values table

The following table holds the values of the lab experiment. The table only shows the cleaned values. For original data, see Appendix [C.](#page-122-0) The data set was created in cooperation with Michael Grömer  $\,$ (Grömer, [2021](#page-87-0)).

- *No*. $=$  Experiment number
- $m_{C,t_1}$  = Mass of the cathode before the experiment
- $m_{A1,t_1}$  = Mass of anode 1 before the experiment
- $m_{A2,t_1}$  = Mass of anode 2 before the experiment

 $m_{C,t_2}$  = Mass of the cathode after the experiment

- $m_{A1,t_2}$  = Mass of anode 1 after the experiment
- $m_{A2,t_2}$  = Mass of anode 2 after the experiment
- *d* $d =$  duration time
- *U* $U = Vol$ tage
- *T* $T = \text{Temperature}$
- $I =$ Amperage



<span id="page-103-0"></span>96



97



98












Table B.<sup>1</sup>.: Documented lab values

The following table holds the calculated values for each experiment based on the original values fromTable [B.](#page-103-0)<sup>1</sup>.

- *No*. $=$  Experiment number
- ~*j* $j =$  Current density
- *A* $=$  Platting area
- $\Delta m_C$  = Cathode mass difference before and after the experiment
- ∆*mA*1 <sup>=</sup> Anode <sup>1</sup> mass difference before and after the experiment
- $\Delta m_{A2}$  = Anode 2 mass difference before and after the experiment
- *te* $e =$  expected layer thickness
- *tr* $\epsilon_r$  = real layer thickness
- ∆*t* $t =$  Deviation of expected to real layer thickness
- *P*= Electric power
- *W*= Electric potential energy



















Table B.<sup>2</sup>.: Calculated lab values

## Appendix C.

## Lab values

The following pages contain the original documentation of the lab experiment.

Versuch #1



























 $447.08$ 44 17.04  $47 16.63$  $A771665$  $48 16.79$ 48 16,82













 $19 \frac{9}{4662}$  $W515$   $S_0Q$   $25m_1$  $20°C$  $1 - \epsilon$  we | Flückel  $ln|_{\infty}$ Vir Charles  $\begin{array}{|c|c|c|}\n\hline\n18,97 \\
12,60 \\
12,95\n\end{array}\n\quad\n\begin{array}{|c|c|}\n\hline\n6,6 & & \\
6,6 & & \\
5,1 & & \\
5,1 & & \\
\hline\n\end{array}$  $18, 17$ Prode 1  $12,74$ Anode 2  $13,09$  $T_{2\mu\rho}$ :  $2(7)$  $T_3: 0,582$  $T_{2}$ : 0,582  $L, 0.597$  $V: o, 8$  $M_{op}$ : 10:28 (25 min)  $M_{\text{out}}$ : 10 = 23  $40^{\circ}$   $24^{\circ}$  $W516$   $$80$   $35$   $\frac{1}{2}$ 14,04 14,82 11 3,6  $\frac{1}{48,49}$  $\begin{array}{c|c} A & 1 & 14 & 6 & 3 \\ A & 2 & 15 & 49 \end{array}$  $T_{\text{supp}} = 39.1$  $\hat{I}_{2}: 7770$  $\hat{1}$  :  $\hat{1}$   $\hat{6}$   $\hat{2}$  $T_{3}$ : 1,764  $V: 0.3$  $Youx: 70:01$  $2.56$   $10:36$ 35 min



 $W8R$  $Soll 65 min 20°0 0,700$  $V_1$  18,24  $V_2$  18,24 18,49  $V_1$  16,38  $V_2$  14,24 16,24 16,24 16,24 16,24 16,24 16,24 10 Flocks 4  $T_1: 0.195$   $T_2: 0.195$   $T_3: 0.195$ Ebrah 11:13 Showheir 11:00 2 45 min 40°C 33 st WS19<br>4 15,76 16,71 4,65 3,45<br>02 16,40 15,88  $Sell$  45 mg Temp: 39,5  $J_3$  1062  $I, 1028$  $I, 1,050$ <br> $V, 1, 1$ Entrain 11:58 45 mil Short rail 11:13



 $10323$  Sell 40 in 60°C  $214/42$  $\begin{array}{|c|c|c|c|}\n\hline & S/\alpha\Lambda & S/\delta\ \hline & 15,61 & 16,46 \\
\hline & 12,59 & 12,31 \\
\hline & 14,00 & 12,31\n\end{array}$ Flüch  $\frac{1}{\theta}$  $\frac{463}{(20,66119)}$  $44,02$  $\theta$  2  $1 - 13, 77$  $[4p 59]$  $V: o.5$  $I_2 0,642$   $I_3 0,641$  $I_1$  0,661  $6 - 42 : 35$ Shal 12:55



 $A = 3.6 - 4.61 - 2$




 $4 = 4.65 - 3.05 - 2$ 



 $1,107$ 



 $F_{xx}$   $T = 16:00$  $-16.20$ 

 $20 - i$ 









 $\mathbb{P}$ 



 $\alpha$ 

 $\overline{\mathbf{w}}$ 







 $\mathcal{D}^{\text{max}}_{\text{max}}$ 



 $\widetilde{\mathbb{R}}$ 



i<br>K



 $\mathcal{A}$  . The set of  $\mathcal{A}$ 



 $\hat{\mathbf{z}}$ 

ź





 $\langle \hat{\mathcal{S}} \rangle$ 





 $\tilde{\mathcal{A}}$ 





 $\overline{\phantom{a}}$ 





















 $\langle \begin{array}{c} \bullet \\ \bullet \end{array} \rangle,$ 

 $\tilde{\Omega}$ 







 $\tilde{\varepsilon}^{\tilde{\omega}^{\prime}}$ 





 $\tilde{\mathcal{L}}$






 $\sim 10^{11}$  km s  $^{-1}$ 

ò,





















è.





 $\alpha$ 





 $\mathbf{w}^{\prime}$ 





 $\widetilde{\sigma}_1^*$ 



 $\bar{u}$ 

 $\frac{\epsilon}{\epsilon}$ 



 $\tilde{\alpha}$ 













 $\label{eq:2.1} \begin{array}{ll} \mathbb{E}[\mathbb{E}$ 











 $\sim$   $\sim$