Konferenzband 2022 I Proceedings 2022

Masterarbeiten I Master Theses

Cold forming Simulation by means of the Finite Element Method

Sebastian Bauer, and Franz Josef Falkner (supervisor)

Abstract—**Cold forming is a widely used forging technique that shapes metal materials at near room temperature and enables to produce large amounts in a fast and cost effective manner. However they introduce residual stresses in the steel components and therefore change the resulting stress range which affects their fatigue life. It is therefore important to consider these influences as early as possible in the product development process, which is why two master thesises at the Mechatronics Department of the MCI dealt with this topic. Therefor two cases of cold forming applications were provided by the company of STIHL Tirol Ag and were examined in this thesis to develop a methodology for the simulation of cold deformation processes by means of the finite element method. Because of the consideration of plasticity and the implementation of contact mechanics, several non linearities were introduced that allow a more realistic representation of the actual forming process but lead to the result that non linear simulations have to be carried out. Therefor the methodology is described by the setup of the two cases, followed by a validation of nonlinear input parameters. Following the simulation results are investigated in the form of stress and strain distribution as well as the force reactions on the forming tools. Results of this thesis show, that the simulation methodology was developed successfully and that there are inhomogenous distributions of residual stresses as well as plastic strains in the parts after the forming processes. Further, quite high plastic strains and resulting force reactions were observed. The plastic strains of the deformed parts are then used further as input data for the second thesis with the title "Fatigue life simulation process for coldformed parts". In addition the force reactions at the forming tools can be used for the optimization of existing forming operations or even for the development of new cold forming mechanisms.**

Index Terms—**Cold forming, FEM Simulation, large strains, material nonlinearities, contact mechanics.**

Sebastian Bauer studies at MCI, Innsbruck, Austria, E-mail: s.bauer.1223@gmail.com

F.-J. Falkner is with the Department of Mechatronics, MCI, Innsbruck, Austria.

I. Introduction

"IRON and steel industry is one of the most energy intensive and largest contributing industries to global greenhouse gas emissions." [1] was the key message of an article in which an overview of emissions caused by the global steel industry in various countries is provided. Therefore, it is recommended to use this material as efficient as possible in the design of new components or machines, which has only become possible by the widespread implementation of simulation software. Amidst the most common simulation methods is the Finite Element Method to determine the deformation and the resulting stress and strain relations during or after loading conditions. Since most of the simulations are used to determine the structural stability of parts, mainly the elastic region is used, because this methodology is sufficient for most engineering applications. For the application to fatigue analysis, more attention has to be drawn to the plastic behavior of metal materials. However, the influence of cold forming processes on the fatigue behavior of steel components is partly unknown. In order to address this complex topic in more detail, two master's thesises are being carried out at the MCI's Faculty of Mechatronics. While the fatigue behavior will be discussed in more detail in further work, this thesis deals with the development of a methodology for the

set up of a cold forming simulation model in ANSYS Workbench.

II. Leading Question

One of the main goals for this thesis is to develop a methodology for the simulation of coldforming processes. For this manner two different cases of cold forming applications are investigated.

A. Case 1: Lawnmower Axle

The first part of interest is the axis of a lawnmower, produced by STIHL. The material is a S235 steel and it is a produced by the bending of a metal wire with a diameter of 12mm, as can be seen in figure 1

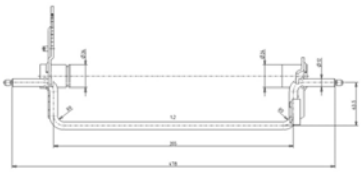


Fig. 1. 2D CAD Draft of the Lawnmower Axle

B. Case 2: Lawnmower Handle

The Handle was produced by the stamping of a Steel Pipe with an outer diameter of 25mm and a wall thickness of 2mm, whereas the same material of S235 stell was used.

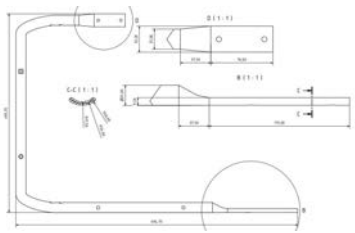


Fig. 2. 2D CAD Draft of the top part of the Lawnmower Handle

For the forming of the handle, detailed pictures were available after a fieldtrip to Farthofer GmbH in Wörgl, Tyrol. The Forming takes place by pressing the stamp (3) onto a raw metal pipe(2), which can be seen in figure 3.

Fig. 3. Industrial Application of the coldforming Process

The pipe was hold in place by a fixing mechanism (1) and a forming support (4).

III. Nonlinear Dynamic Finite Element Analysis

In a linear FEM analysis, the effects of a certain load on a component are investigated, whereby stresses and strains can be calculated directly by a linear correlation between stress and strain. The linear relation is often limited for practical applications due to many factors like the following that lead to a non linear system according to [2]:

- non -linear material law
- geometric non- linearity
- contact mechanics

Especially the material non linearities are considered for the representation of the plastic material behavior.

A. Plasticity Models

Basically the models for elastic-plastic behavior start with a decomposition of the total strain into an elastic and a plastic part. For both parts separate models are implemented. The plastic part can be described by three essential characteristics of the

plasticity models according to [3], namely the flow rule, the yield criterion and the hardening rule which is explained in more detail in the following section.

B. Isotropic Hardening

For many materials, the yield criterion is defined by the evolution of plastic strain. This correlation is called hardening and is defined by the hardening rule. After continuing loading after the yield strength is used, the stress as well as the plastic strain increase even further for a material that has elastic-plastic behavior. This hardening behavior can be described by two common hardening rules according to [3]. For isotropic hardening, plastic loading increases the yield surface, and therefore the yield stress in uniform direction. This type of hardening is used in order to model the behavior of materials under monotonic loading and elastic unloading. In contrast to isotropic hardening, kinematic hardening does not lead to an uniform increase but to a shift of the yield surface. Further, kinematic loading is observed for cyclic loading of metals which is not used for this application [4].

The hardening behavior of metals can be described by the approximation of different material models. In this thesis the application of the multilinear hardening models is investigated. As the name implies, multilinear isotropic hardening is described by the combination of multiple linear stress- strain curves with different slopes, as can be seen in below figure 4.

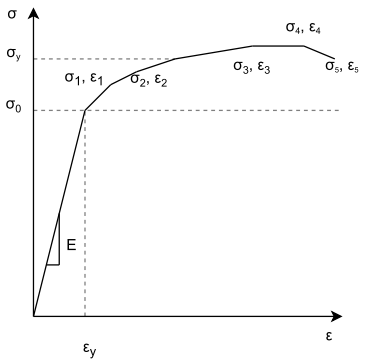


Fig. 4. Multilinear stress strain diagram after [3]

This model is constructed by a piece-wise linear stress-strain curve. Whereas the start of curve is at the origin and the first stress-strain point is correspondent to the yield stress where the following data points define the elastic-plastic material response.

C. Material Properties

Since the success of the simulations depends highly on the used material type, it is a very important topic that has to be addressed further. The material which was used by STIHL Tirol for the manufacturing of the parts was a steel of the category of with a yield strength of 235 MPa. For the implementation to the simulation a very similar steel, namely S235 was used, where the ”S” indicates the application in the field of steel construction. Although this material may have different characteristics regarding the direct manufacturing process, it can be used for the development of a simulation methodology.

IV. Simulation Setup

For the simplification of the real world system it is very important to know the real application in detail, in order to keep the simulation model as simple as possible. This is achieved by neglecting unnecessary details, removing sharp edges or taking advantage of

the symmetry, which is conducted in the following sections.

A. Model Setup Case 1

The final simulation elements for the axle are exposed in figures 5 and 6.

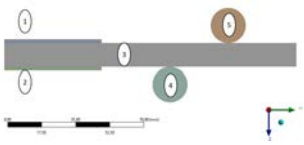


Fig. 5. Side view of the element geometry of case 1

Fig. 6. Front view of the element geometry of case 1

The model consists of five bodies that represent different functions during the forming process. The two rectangular elements (1) and (2) represent the clamping of the axle(3), whereas the upper round element represents the forming mandrel (5) and the lower one represents the support mandrel (5).

B. Model Setup Case 2

In comparison with the first case, the simulation of the handle represents a more complex example, which is why it was carried out after the first case yielded sufficient results. The final simulation setup is depicted in figure 7.

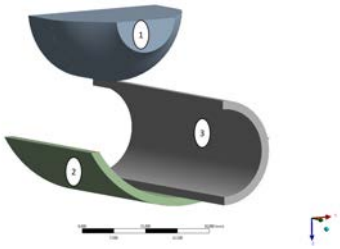


Fig. 7. Isometric view of element geometry of case 2

In above figure the ”C” shaped body in the middle represents the raw pipe (3), which had to be defined as ”flexible” in the program settings. The lower body represents the forming support (2) and the upper body represents the forming stamp (1).

C. Mesh Generation of Case 1

At first the complete body of the axle was meshed to generate 3D solid elements all across the body. For the rigid bodies only a face meshing had to be conducted on selected faces at the surface. The result of this process is presented in figure 8 with a very small mesh size of 0,5mm.

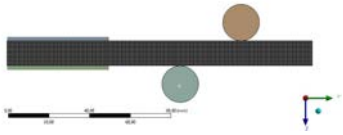


Fig. 8. Mesh generation on the axle

It can be seen, that there is a very well structured and well sized mesh all across the body of the axle. However, if one takes a closer look at the cross-section of the axle, as shown in figure 9, this impression changes.

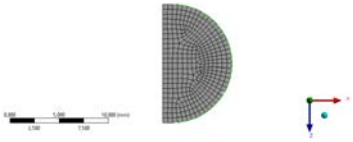


Fig. 9. Mesh generation on the cross section of the axle

D. Mesh Generation of Case 2

The pipe was meshed according to the first case with volume elements and face meshing for the contact surfaces. In below figure 10 a 3D view of the assembly is shown with a mesh size of 1mm whereas

the later simulation was conducted with a mesh size of 0.5mm.

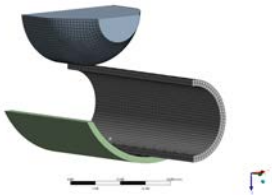


Fig. 10. Mesh generation for case 2

Similar to section IV-C, the mesh looks sufficient for the surface as well as the cross section resolves in a regular mesh.

E. Boundary Conditions for Case 1

Before the load conditions can be defined, it is beneficial to deal with the support of the individual bodies first, where the supports as well as the support mandrel was fixed by a remote displacement with constraints in all degrees of freedom. The axle itself was constrained by the friction between the two supports and a remote displacement that constrained the movement only in Y-direction. This was conducted in order to prevent a rigid body motion of the axle during the first steps of the simulation. Following the loading can be defined, which was achieved in this case by a remote displacement, which can is shown in figure 11.

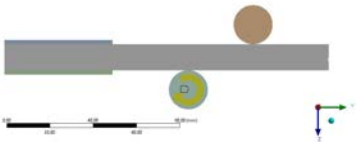


Fig. 11. Boundary conditions on the axle

It can be seen, that the forming mandrel rotates around the support mandrel clockwise, following a defined pattern which can be seen in table I.

TABLE I
BOUNDARY CONDITIONS: LOAD STEPS FOR CASE 1

step number	**simulation time**	**rot. displacement**
	s	°
1	0	20
2	1	80
3	2	91
4	3	88

In order to increase convergence behavior, the loading is conducted in smaller steps at first to 20° and 80° until the maximum at 91° is reached. Due to the elastic portion of the strain, the process has to achieve a small over bend, which is why the maximum rotation is one degree further than the desired bend at 90°. Because only the stresses that were introduced into the material during the bending operation are of interest, a backwards rotation to 88° is realized to relax the part.

F. Boundary Conditions for Case 2

Similar to previous section the support was fixed by a remote displacement without any degree of freedom. Similar to previous section, the raw body has to be constrained into Y- direction, which is feasible due to the real manufacturing process. Since there are two end parts of the metal pipe, they are formed simultaneously and thereby locking each other in translational direction along the tube. Hence the forming stamp is the only body that moves due to a remote displacement in a translational manner, which is exposed in figure 12.

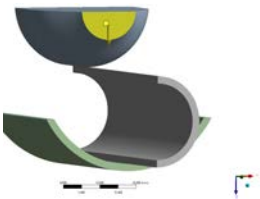


Fig. 12. Boundary conditions for the handle

Since the rigid body motion is not that complex compared to the first case, the loading can be achieved by just two steps that are listed in table II.

TABLE II
BOUNDARY CONDITIONS: LOAD STEPS CASE 2

step number	**simulation time**	**translat. displacement**
	s	mm
1	0	21
2	1	20

It can be seen, that the stamp descends 21mm into positive Z direction and realizes the deformation of the raw pipe. As soon as the stamp has reached its final position it retrieves again around 1mm in order to counteract the elasticity.

G. Contact Definition

In general, ANSYS Mechanical supports two different types of contact relationships. The first type is the connection between two solid bodies, where the second one describes the connection between a flexible and a rigid body. Although the forming tools have a certain flexibility in reality, they can be considered as solids in the simulations because their stiffness is considerably higher than that of the raw piece. [5]

Further there are many more options provided by the contact settings of ANSYS. While most of them can remain unchanged a few additional settings have to be done which are exposed in table III.

TABLE III
ADDITIONAL CONTACT SETTINGS

definition	
type	frictional
frictional coefficent	0, 18
behavior	asymmetric
advanced	
formulation	normal lagrange
geometric modification	
interface treatment	adjust to touch

In the first section the contact is defined by its type, which is a frictional contact (μ=0,18), and its behavior. In reality the contact between forming tools and the raw parts would be considered as symmetric because the load operates into both contact surfaces. However the hardness and stiffness of forming tools is significantly higher than of the raw part which is why the contact behavior can be modeled as asymmetric.

For the first case, all of the contact surfaces are located at the touching points of the tools on the axle. This is also the case for the second case, whereas an additional contact is created during the course of the simulation as soon as the inside of the tube gets into self contact.

H. General Analysis Settings

In order to conduct a nonlinear simulation, the analysis settings have to be adjusted according to table IV.

TABLE IV
ANALYSIS SETTINGS FOR NONLINEAR ANALYSIS

step controls	
auto time stepping	On
define by	substeps
initial substeps	30
minimum substeps	30
maximum substeps	1000
solver controls	
weak springs	On
large deflection	On

As soon as the Auto Time Stepping option is enabled, a nonlinear simulation process is conducted. In order to increase the convergence behavior sub steps should be used which were defined between 30 and 1000 sub steps. In order to simulate the strong deformations during a coldforming analysis, the options of ”Weak Springs” and ”Large Deformation” have to be enabled.

V. Results and Interpretation

In the course of this thesis, simulation results for equivalent stress, equivalent elastic strain and equivalent plastic strain were determined. However, for the later application in the fatigue analysis, only the equivalent plastic strains are of further importance. Therefore those results will be discussed in the following sections.

A. Equivalent Plastic Strain for Case 1

Following to the distribution of stress in the previous section, the highest values for plastic strain will also be seen at the contact points with the mandrels, which can be seen quite clearly in figure 13.

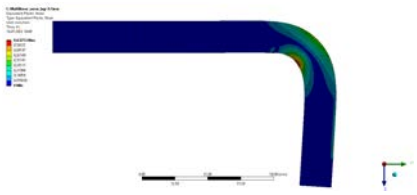


Fig. 13. Equivalent plastic strain distribution for case 1

Furthermore, it can be assumed that the stresses and the resulting strains along the neutral fiber approach zero or at least have their minima, which can be seen in more detail in figure 14.

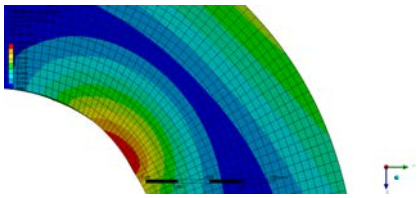


Fig. 14. Equivalent plastic strain over cross section for case 1

It can be seen that the maximum value occurs exactly in the middle of the bend at the inner side. Theoretically, the maxima should appear symmetrically at the innermost as well as on the outermost fibers. This implies, that the elements inside the neutral fiber experience much higher loading than the elements laying on the outer side. One reason for this phenomenon may be the superposition of different load conditions.

B. Interpretation of Case 1

The maximum penetration is set into relation to the geometrical dimensions which is described in this case by the diameter of the axle. According to [6] the relation between the maximum penetration and the geometric dimension should be lower than 1%. With a value of 0,199 % this is the case, which is why one can assume that the penetration has a minor influence on the simulation results. The distribution pattern of the equivalent plastic strain looks quite symmetric, as can be seen in figures 13 whereas the loading increases in the direction to the inner radius as can be seen in figure 14. The maximum plastic strain is at 63,25 %. In addition to the elongation strain that is achieved by the deformation operation, the strain that is caused by the contact at the support mandrel leads to a superposition of loading conditions. A further influence by the frictional stress in between is supposed which results in the end to such a high equivalent plastic strain.

C. Equivalent Plastic Strain for Case 2

After the equivalent stress distribution and the shape of the deformed part was discussed previously, the distribution of equivalent plastic strain is addressed here. In figure 15 the plastic strains which were introduced into the material during the forming process are exposed across the whole part.

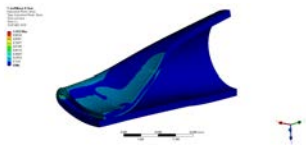


Fig. 15. Equivalent plastic strain distribution at the handle

The extremely high equivalent plastic strains, ranging from the maximum at 111% to around 61 %, only occur locally in the area where the material experiences the highest deformation.

VI. Interpretation of Case 2

Since the contact surfaces have a significantly higher influence at the simulation result for the second case than for the first case, the contact parameters have to be observed carefully. Because the relation of the maximum penetration over the complete forming process with respect to the wall thickness of the tube is only at 0,071%, it is far lower than the treshold of 1 %. Therefore it can be stated, that the penetration has a minor influence on the simulation results. Similar to the first case, the maximum equivalent plastic deformation is also found at the place of highest deformation with a value of 110,5%, which is why a superposition of load conditions has to be assumed as well.

VII. Conclusion and Outlook

The goal in this work was the establishment of a methodology in order to simulate the behavior of steel during cold deformation. This goal was achieved by means of a minimum working example and could successfully be applied to a more complex model. In addition the reaction loads were determined that can be used for the further development of forming tools and mechanisms. To put this methodology into direct industrial application, a more in depth look at the real process parameters has to be conducted. Further a tensile test with the used material would increase the accuracy of the material model. Based on the results of this thesis, a second master thesis with the topic ”Fatigue life simulation process for cold formed parts” was executed by Mr. Raphael Sigloch at the MCIs mechatronics department.

Acknowledgment

Throughout the work of this thesis, I was very glad to receive a lot of support and assistance by many persons. First of all I want to thank my supervisor, Dr. techn. Franz-Josef Falkner for his patience and support. My fellow student colleagues I want to thank for their support and motivation over the whole course of this thesis. Finally I want to thank my parents for the financial and emotional support during my whole studies and especially during the masters program.

References

[1] A. Hasanbeigi, “Global steel industry’s ghg emissions,” https://www.globalefficiencyintel.com/new-blog/2021/global-steel-industrys-ghg-emissions, 2022, global Efficiency Intelligence.

[2] C. Gebhardt, *Praxisbuch FEM mit ANSYS Workbench*, 2nd ed. Hanser, 2011.

[3] “Material reference,” PDF, 2022, aNSYS, Inc. and ANSYS Europe.

[4] M. Hofer, “Fatigue strength,” PDF, 2020, lecture Notes - MCI.

[5] “Ansys mechanical advanced connections,” PDF, 2017, aNSYS, Inc. and ANSYS Europe.

[6] P. Wriggers, *Computational Contact Mechanics*, 2nd ed. Springer, 2006.

Sebastian Bauer graduated in mechanical engineering in 2020 at the University of applied sciences in Rosenheim. Following he moved to Innsbruck, Austria to begin his masters degree course in mechatronics at MCI, e-mail: s.bauer.1223@gmail.com

Assembly of an optical characterization bench dedicated to the coupling between a laser and an optic photonic chip (OPA)

Germain Favier, Sylvain Guerber (CEA-Leti supervisor), and Sebastian Repetzki (MCI supervisor)

***Abstract*—Light imaging detection and ranging (LiDAR) is the light equivalent of radar technology and is a method suitable for self-driving cars obstacle detection. Using an optical phased array (OPA), a small photonic chip which can steer the laser beam in different directions, the LiDAR system could be implemented in cars with low cost and high precision.**

To find a suitable laser and OPA combination for self-driving cars, a characterization bench is build to determine the interaction characteristics between the OPAs and tailor-made lasers. The lasers that will be used on the bench emit an elliptic and divergent beams. So the bench needs to position the laser source edge to edge with the edge coupler of the OPA, while allowing for control of the roll, yaw and pitch of the laser source and giving reliable and repeatable measurements.

The bench uses a camera and a microscope to monitor the position of the laser source respective to the edge coupler of the OPA, and uses image processing on the obtained images to measure the position and orientation of the laser source. A parallel robot (Steward platform) is put on the bench to give the laser source six degrees of freedom, and a piezoelectric linear actuator (nanocube) allows the laser to make small and repeated translations in the three directions in space. An algorithm was developed to determine the position of the laser source in 3D, and Python classes were created to use the bench and its different elements.

The classes created to manipulate the bench allow the characterization of a laser source and a photonic chip within one hour, starting from scratch on the bench. The optimal position, with the maximum optical power transmitted, of the coupling can be determined clearly for a laser source and photonic chip. Using the algorithm designed for the bench, the position of the laser source could be determined at a 10 µm precision on the X-, Y- axis of the hexapod and at a µm precision on the hexapod Z-axis. The accuracy in the found position of the laser source allows the control of the orientation of the laser with almost no translation. Lastly, a contact detection method was deployed on the hexapod to know when the laser source comes in contact with an obstacle.

The bench which has been build and tested, will be used to characterize photonic chips with elliptic and divergent laser sources, in particular OPAs. The characterization bench will be used in European project VIZTA to study the use of OPA for possible LiDAR applications in self-driving cars.

***Index Terms*—Stewart platform, Piezoelectric linear actuator, Image processing, Characterization Bench, OPA, VIZTA**

I. Introduction

TO develop widespread self-driving cars, a reliable and low-cost 3D imaging technique is needed. One such 3D imaging technology is light detection and ranging (LiDAR), where the distance between the emitter and any obstacle can be measured through the time of flight (ToF) principle. ToF works by knowing the speed of the wave sent and measuring the time the wave takes to come back. Doing so the distance to any obstacle can be measured. To steer the optical beam in the LiDAR and determine the distance to obstacles in 3D from the emitter, an optical phased array (OPA) can be used. An OPA use the constructive and destructive interaction of the light to steer the light beam in one direction, by having constructive interference of the laser light in the desired steering direction and destructive interference in all other directions. Current estimates indicate that a LiDAR used on self-driving car needs to detect an obstacle from at least 200 m [1] among other specifications. A characterization bench needs to be build to determine the characteristics (optical power, angle resolution and more) that can be transferred from different laser sources to different OPAs output, and to find which laser source and OPA combinations are suitable for application of LiDAR in self driving cars.

Characterization benches and the process by which a laser source and an OPA are characterized already exist [2]. But the bench build here varies in the type of laser source and the coupling technique used. The coupling technique used in [2] is a grating coupler. Inside a photonic chip, such as OPA, waveguides are used to guide the light. Grating coupling function by sending the light at an angle from the waveguide as shown in figure 1. Grating coupling makes the alignment of the laser source and the edge coupler of the OPA permissive to error, but at the price high optical power losses [3]. Instead an edge coupling technique is used on the bench presented here. The light is send in the direction of the photonic chip waveguide as shown in figure 2. Edge coupling exhibits less losses and a higher bandwidth than grating coupling, but is more sensitive to alignment error [3].

G. Favier, Management Center Innsbruck, e-mail: fe7337@mci4me.at
S. Repetzki, Department of Mechatronics, MCI, Innsbruck, Austria.

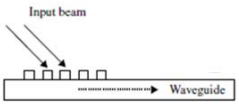


Figure 1. Grating coupling as presented in [4]

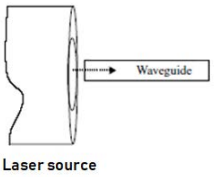


Figure 2. Edge or butt coupling as presented in [4]

The laser sources that have to be characterized on this bench send an elliptical and divergent beam, making the edge coupling even more difficult. The edge coupler of the OPA has to have an elliptic profile as well to get as much power as possible from the laser source. These 2 elliptical profiles have to be superposed by the bench. To do so, the laser will be controlled by a 6 degrees of freedom mechanism to control the orientation and the position of the laser source. Since the laser beam is divergent the laser source and edge coupler will have to be put as close as possible to each other during characterization. Finally, the mechanism will need to test quickly many position where the coupling might happen to find the best coupling position where the maximum optical power is transmitted. The different parts of the characterization bench and the procedure to make a fast and reliable coupling of a laser source and a photonic chip is detailed in the following sections.

II. Methods

A. Bench set up

To manipulate the laser source, it is put on an actuator using piezoelectricity [5] called Nanocube which itself is put on a hexapod [6][7] both made by the company physic instrument (PI). The nanocube from PI is composed of an immobile part and a mobile part, actuated by 3 piezoelectric elements perpendicular to each other which correspond to the 3 axes seen in figure 3. The nanocube movement are controlled through a PID controller by applying a voltage proportional to the desired position on each of the axis. With the nanocube, high precision and speeds can be achieved. The nanocube is used to make the fine tuning of the characterization, after an approximate position is found with the hexapod.

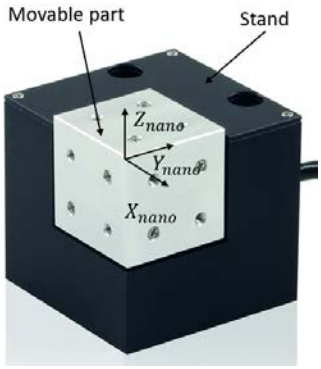


Figure 3. Nanocube P611 from PI (Physik Instrumente)

The Hexapod is a parallel robot composed of an immobile platform called Base and a mobile platform called Platform, linked by 6 arms which are linear actuators as seen in figure 4. The elongation of the arms determine the mobile platform position and orientation. The hexapod used on the bench can use different coordinates systems created by the user starting from the original coordinate system, called Hx, positioned at the center of the mobile platform when all the arms are at mid-length. 2 coordinates systems are of particular importance, the Worskpace, abbreviated W in schematic, and the Tool, abbreviated t in schematic. Both allow for rotation around their origin, but the Worskpace origin is fixed in space while the Tool follows the movements of the hexapod. So the Tool can be attached to a real object, in our case the laser source we want to manipulate. The workspace origin is put at the laser source position when the hexapod is at its origin to measure the laser source translation from this reference position. With the Tool, the laser source roll, yaw and pitch can be corrected without translation.

Figure 4. Hexapod H811 from PI (Physik Instrumente)

To monitor the alignment of the laser source and the photonic chip, the characterization bench will use a camera and a microscope. The camera and microscope are fixed on a stand with 2 perpendicular manual actuators to move the field of view of the camera on the bench. The photonic chip

is installed on a stand composed of 2 elements that can be translated or rotated manually. The different elements of the bench will be installed so that the coordinate systems of each elements of the bench are aligned on the camera frame. So the X- and Y-axis of each coordinate system will be aligned with the h and v axis of the camera field of view during the bench set-up. Lastly, to measure the optical power that is emitted by the photonic chip, an optical power meter is present. The elements of the bench and their layout are depicted in figure 5, depicted with the desired position of the tool coordinate system and the workspace coordinate system.

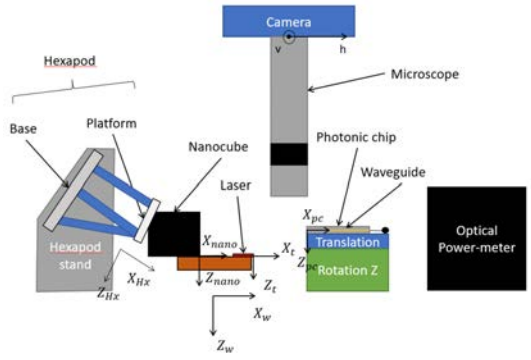


Figure 5. Heat sink and stand

The bench will be piloted with a Python program developed using object-oriented programming (OOP) to follow the bench changes (if actuator are added or removed). The user will interact with the class characterization bench to pilot the characterization bench. The class characterization bench contains one instance of camera, nanocube and hexapod. The summary of the classes interaction and their different functions can be seen in figure 6

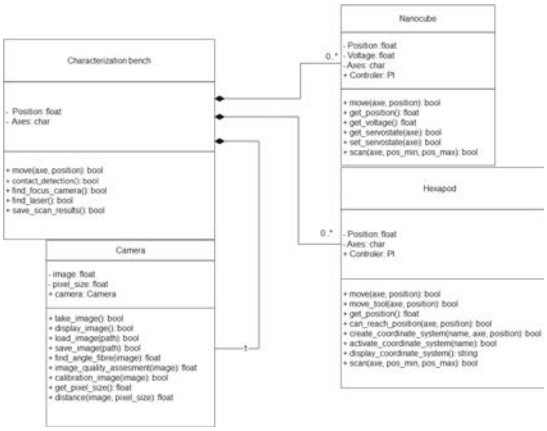


Figure 6. UML diagram of the characterization bench

The following Python libraries are used: NumPy, scikit-image, OpenCV and the PI Python library to manipulate hexapod and nanocube.

B. Laser and photonic chip alignment

The first step to characterize the laser source and photonic chip is to align them both on the axis perpendicular to the camera plane. The depth of field of the microscope used on the bench is small (3.5 µm). So the image is blurry outside a small height range. First, the microscope is focused on the photonic chip to clearly see the latter with the camera. Then the camera and microscope are moved over the laser source. By moving the laser source perpendicularly to the camera filed of view, and assessing the blur on the image taken with the camera of the photonic chip, it can be determined if the laser source and photonic chip aligned or not. To asses automatically the blur on an image, 2 image quality assessment (IQA) method are used. The first method uses entropy [8] and the second method uses the Fourier transform [9].
To compare both method, a video was taken of the laser source moving from its original position to 5 mm on the Z-axis. Each frame was then asses with both IQA, and the maximum focus found for both IQA were compared, as well as the time it took the IQA to compute them.

Algorithm for IQA with entropy

Step 1: Consider $I_{T(u,v)}$ as the acquired image of size N × N.

Step 2: Normalize the image $I_T(u,v)$:

$\widetilde{I_T}(u,v) = \frac{I_T(u,v)}{\sqrt{\sum_{u,v=0}^{N-1}(I_T(u,v))^2}}$ such that $\sum_{u,v=0}^{N-1}\left(\widetilde{I_T}(u,v)\right)^2 = 1$

Step 3: Calculate the probability of an image $\widetilde{I_T}(u,v)$: $p(r) = \frac{n_r}{N \times N'}$

where r is the gray levels and n_r is the total number of pixels in each gray level

Step 4: Calculate the proposed image focus measure:

$I_\tau = \log_\alpha m - \log_\alpha m \sum_{j=1}^{n} p_j^\alpha$ |

Step 5: If $I_\tau \approx 1$, good quality image, else degraded image

Figure 7. Algorithm for IQA from [8]

Algorithm for IQA in Frequency Domain

Step 1: Consider Im as the acquired image of size $M \times N$.

Step 2: Compute F which is the Fourier Transform representation of image Im

Step 3: Find F_c which is obtained by shifting the origin of F to centre

Step 4: Calculate $AF = abs(F_c)$ where AF is the absolute value of the centered Fourier transform of image Im.

Step 5: Calculate $Max = \max(AF)$ where Max is the maximum value of the frequency component in F.

Step 6: Calculate T_H the total number of pixels in F whose pixel value > $threshold$, where $thres = \frac{Max}{1000}$

Step 7: Calculate Image Quality IQA with the formula $IQA = \frac{T_H}{M \times N}$

Figure 8. Algorithm for IQA from [9]

From the Fourier and the entropy algorithm, a maximum is expected when the image is focused. Their complexity are respectively $O(nlog(n))$ and $O(log_2(n))$ with n their number of pixel.

C. Laser source and photonic chip orientation

To make the laser source face the edge coupler, the angle of the laser source needs to be determined. The stand on which the laser is installed on the nanocube and hexapod is supposed to be straight and so it is assumed that the angle around X_W- and Y_W-axes don't have to be corrected. To measure the angle on the Z_W-axis, 2 step are realized. In the first step, the edges on the image are found with the Canny edge filter [10]. The edges found should match the edges of the laser source and are assumed to be straight lines. Then the angle of the edges are found with the Hough Transform [11]. Once the edges angles are found, the average of the angles is computed to estimate the angle to be corrected by the hexapod.

D. Laser source position in 3D

With the camera, a calibration is realized by measuring a known distance between 2 points to convert a distance from pixels to mm. To measure the position of the laser source from the hexapod origin and use the Tool coordinate system to change the orientation without translation of the laser source, a method is designed to measure the distance from the laser source to the center of rotation. A first approximation of the laser source position by measuring the distance between the center of rotation of the hexapod (center of the mobile platform) to the laser source is realized with a CAD software. Then the method detailed here will allow to measure the distance between the center of rotation and the laser source. The process is design to be iterative, where the user ask for a precision and the process can be repeated until achieving the desired precision. In figure 9, the translation from the laser source in position 1 LS_1 to position 2 LS_2 after a rotation of α is depicted. The goal is to find the distance d_{tool} in the W coordinate system between the laser source and the Tool position to correct the position of the latter on the hexapod. On the bench, the user asks for a rotation α around the Z_t axis, and measures the translation of the laser source with the camera.

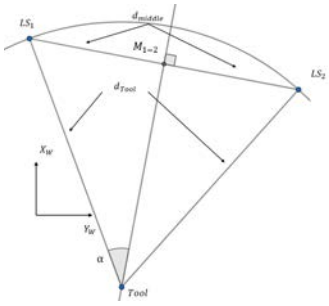


Figure 9. Distance between Tool and laser source in X-Y plane from workspace

So the distance between the laser source and the Tool d_{tool} can be calculated from the distance $\overrightarrow{LS_1LS_2}$ and the angle of rotation α as:

$$d_{tool} = \frac{||\overrightarrow{LS_1LS_2}||}{2\sin\alpha} \tag{1}$$

The sign of the correction can be determined from the distance sign and the rotation direction. Now to determine the direction in which the tool as to be moved to reach the laser source, a first method was explored. It was assumed that the tool was on the laser body, so by knowing the angle of the laser source as determined previously in sub-section II-C, the distance on the axes X_w and Y_w could be determined. But this approach showed high error, after correction of the Tool position (8 mm), by repeating the measurement of the distance, the laser source was still found to be at 3 mm. Instead it is assumed that the tool is at a distance r on the Y_t axis of the laser body, the laser is put at an angle β from Y_t and that the distance d correspond to the distance between the laser source and intersection of the Y_t-axis and the laser body. The laser position LS is depicted in figure 10.

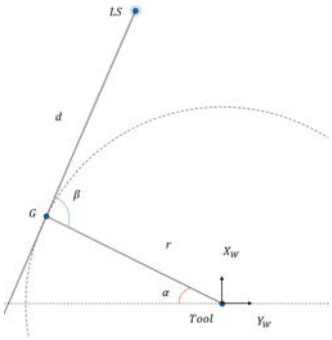


Figure 10. Schematic of distance between laser source and Tool

From the figure, the position of the laser source on the axes X_W and Y_W can be determined as:

$$x = r\sin\alpha + d\sin(\beta - \alpha) \quad y = -r\cos\alpha + d\cos(\beta - \alpha) \tag{2}$$

To measure the translation with the camera and microscope, the rotation angle α around Z_w needs to be small and can be neglected, $\beta = \frac{\pi}{2}$ if the laser source is correctly installed, and r is assumed to be negligible in front of d. Equation 2 becomes:

$$x = d \quad y = 0 \tag{3}$$

And the distance can be found with equation 1. After the distance d is corrected, d becomes negligible before r, and all the previous assumption stay true. Equation 1 becomes:

$$x = 0 \quad y = -r \tag{4}$$

And the distance can again be found with equation 1. This process can be repeated as many time as wished until the desired precision is achieved.

To determine the position on the Z_w axis from center of

rotation is assumed that after the precision is achieved on the X_w and Y_w, the laser source is aligned on the Z_w-axis with the Tool. Now to measure the distance between the tool and the laser source, the laser source is rotated on the X_t-axis from the previous position to α and $-\alpha$ as depicted in figure 11. The distance on the Z_W-axis after the rotation is the same as $\cos(\alpha) = \cos(-\alpha)$. So the distance on the Z_W-axis will create a translation on the X_W- and Y_W-axes which can be measured with the camera. From the translation the distance d_{tool} from the laser source can be determined with the formula 1.

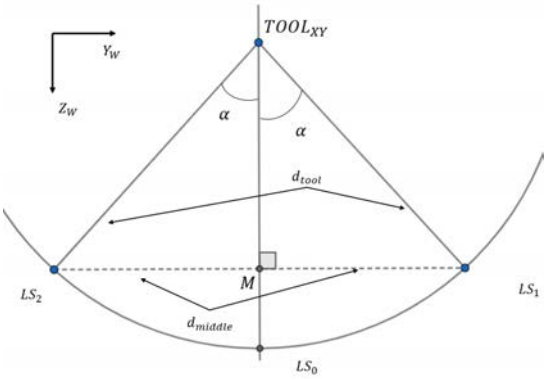


Figure 11. Distance between Tool and laser source in Y-Z plane from workspace

E. Contact detection

To detect a contact of the laser source and the photonic chip, the piezoelectricity of the Nanocube is used. When the laser source encounter an obstacle during its motion, a force F is applied to the nanocube against its direction of motion following the formula:

$$\int_{p_1}^{p_2} F x dx = \int_{v_1}^{v_2} \frac{1}{2} m v^2 dv \tag{5}$$

The article [12] show a piezo-motor with a self-sensing contact detection using impedance. But the nanocube from PI doesn't allow for the measure of the impedance. Instead the variation of the voltage applied by the PID controller to the nanocube to keep the position constant will be measured. With a pulley system the sensitivity of the nanocube's voltage to a force can be determined.The voltage sensitivity determined for the initial nanocube setting on its X-axis is:

$$\frac{\Delta V}{\Delta F} = -0.000\,82\,\mathrm{V/mN} \tag{6}$$

The average and standard deviation of the nanocube with and without force applied to it are measured as well:

axis	X
Average without force[V]	-9,0067
Standard-deviation without force [V]	0,0074
Average with 97.36 mN [V]	-8,9271
Standard-deviation with 97.36 mN [V]	0,0031

Table I
AVERAGE AND STANDARD-DEVIATION OF NANOCUBE

Using the average and standard deviation determined in table II-E, the normal distribution law of the nanocube's voltage can be determined and used to predict the voltage probability depending on the applied force. Then a voltage threshold is determined to differentiate between no forces applied and a force of 2 grams applied to the laser source, and so to the nanocube. The voltage threshold taken on the bench is 0.15 Volt.
The voltage variation of the nanocube is measured with the formula:

$$V_{err} = \sum_{i=1}^{i=3} V_i - Average_i \tag{7}$$

With V_i the current voltage of the nanocube on axis i, and $Average_i$ the average on the i-axis until the current Voltage. Then the contact detection can be tested, the flowchart in figure 12 describe the algorithm to know if there is a contact during motion or not.

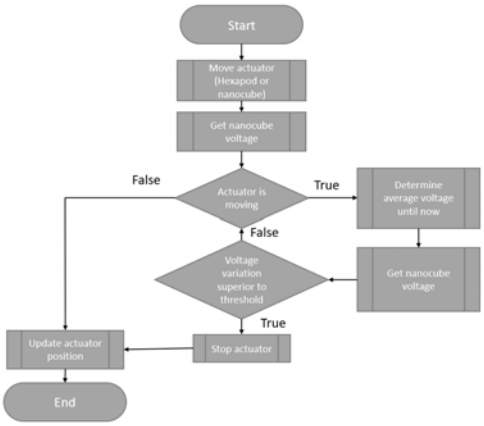


Figure 12. Flowchart of contact detection

F. Bench manipulation and characterization

After the Bench set-up and the laser source and photonic chip are installed on the bench, the characterization can start. The IQA is used to align the photonic chip and the laser source, then the orientation of the laser source is measured and the position of the laser source from the hexapod origin is measured. The workspace and tool coordinate system are put on the laser source when all the linear actuator are at mid length. The hexapod is then used to put the laser source face to the photonic chip while detecting contact with the nanocube, and the orientation measured previously is corrected without translation with the tool coordinate system of the hexapod.

The nanocube is then used to find the position where optical power is transferred and which position as the best output power for the photonic chip. From the coordiante system used on the bench point of view, this would correspond when the Tool coordinate system put on the laser source is superposed with the photonic chip (pc) coordinate system from figure 5. The relation between the different coordinate system until the coupling is summarized in figure 13

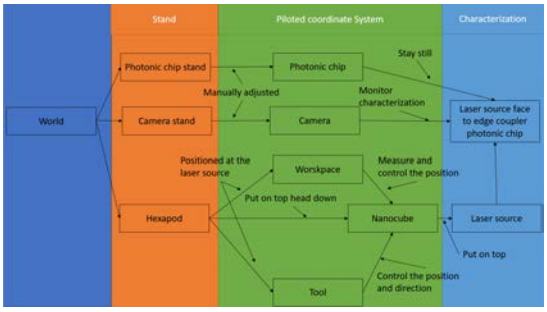


Figure 13. Summary of coordinate system

III. Results

The test of the characterization bench were realized with an optical fiber as laser source, without divergent and elliptic beam, and a photonic chip which was not an OPA.

A. IQA

The clarity of the image compared to the hexapod position fro the Fourier IQA can be seen in figure 14. The frame with the maximum focus according to the IQA is shown in figure 15.

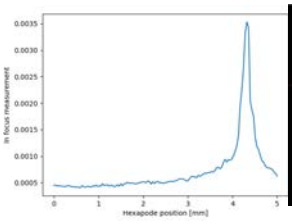


Figure 14. Focus against position with Fourier IQA from [9]

Figure 15. Frame with maximum focus found

The clarity of the image compared to the hexapod position for the entropy IQA can be seen in figure 16. The frame with the maximum clarity is shown in figure 17.

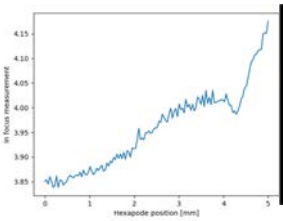


Figure 16. Focus against position with entropy IQA from [8]

Figure 17. Frame with maximum focus found

The time it takes each IQA to compute the quality of the image depending on the resolution is shown in table II.

size of the image	time with IQA fourrier (s)	time with IQA enthalpy (s)
768*1024	0.169	0.053
2048*536	0.356	0.176

Table II
TIME TO COMPUTE IQA

B. orientation

The picture of the optical fiber used as laser source is shown in figure 18.

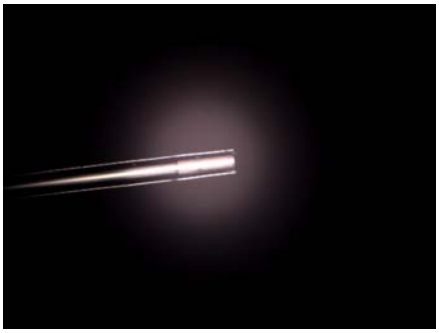


Figure 18. Image of the laser source

The edges found on the image are shown in figure 19, and the corresponding lines are shown in red in figure 20

Figure 19. Edges of the laser source image

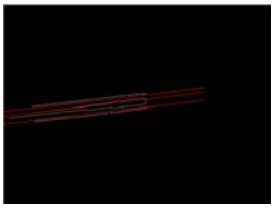


Figure 20. Line from the edges

The sensitivity of the angle found depending on the Gaussian filter size and the step angle are summarized in tables:

size Gaussian filter (σ)	Angle step (°)	Angle found (°)
2	0.1	-3.98
2	0.01	-3.960
5	1	-3.9
5	0.1	-3.75
5	0.01	-3.679
8	1	-3.3
8	0.1	-3.81
8	0.01	-3.719
10	0.01	No angle found

Table III
ANGLE FOUND VARIATION DEPENDING ON ANGLE STEP AND GAUSSIAN FILTER

The time it takes the algorithm to find the angle depending on the angle step, the angles search and the size of the Gaussian filter are summarized in table IV for angles between -21 and 21°, and in table V for angles between 0 and 180°

size Gaussian filter (σ)	Angle step (degree)	time to compute (sec)
5	0.1	0.972
5	0.01	1.874
5	0.001	12.638
8	0.01	1.935

Table IV
TIME TO FIND ANGLE BETWEEN -21 AND 21 DEGREES

size Gaussian filter (σ)	Angle step (degree)	time to compute (sec)
5	0.1	1.314
5	0.01	6.347
5	0.001	130.073

Table V
TIME TO FIND ANGLE BETWEEN 0 AND 180 DEGREES

C. Laser source position in 3D

The laser source distance to the tool is measured by repeating the process once more after the Tool position is estimated. The position as been corrected once on the X-, Y- and Z-axes. The table VI summarize the translation and the distance found after a rotation of 5° on the corresponding axis.

rotation axis	translation on camera [pixel]	d_{tool} from equation 1 [µm]
Z	[4.02, 0.93]	47.22
X	[0.38, 0.30]	5.46

Table VI
DISTANCE FOUND DEPENDING ON THE AXIS

For the same laser source and tool configuration, the sensitivity of the distance found in pixel depending on the rotation angle is shown in table VII.

angle [°]	0.1	0.2	0.3	0.5	0.6
distance [pixel]	6828.17	6830.49	6832.12	6822.42	6815.37

Table VII
DISTANCE FOUND VARIATION WITH ANGLES

D. Contact detection

The voltage variation measured by formula 7 during motion of the hexapod between 0 and 20 mm with a obstacle at 17 mm can be seen in figure 21.

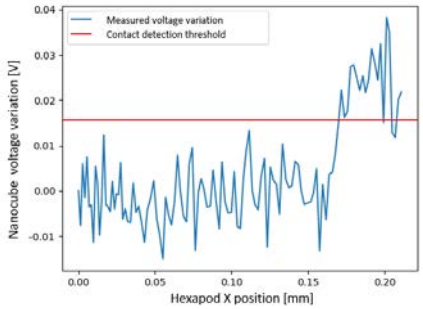


Figure 21. Voltage variation of nanocube during hexapod motion with contact

The voltage during motion of the nanocube with a contact at 25 µm and without contact can be seen in figure 22

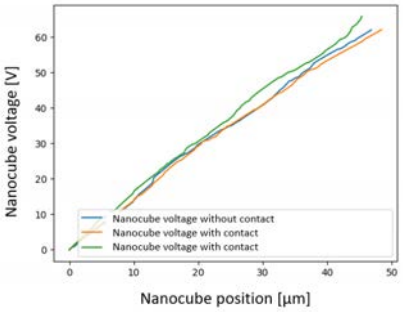


Figure 22. Voltage variation of nanocube during motion

E. The characterization

After the tool is put in front of the edge coupler of the photonic chip, a scan is realized with the nanocube. The result is shown in figure 23, where the output is measured by an optical power meter,and the laser position is tested by moving the nanocube on Y and Z axes.

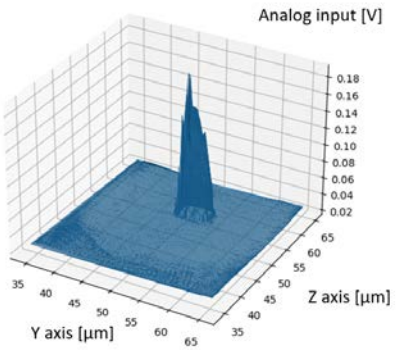


Figure 23. Scan realized by Nanocube

The whole characterization happened in one hour starting from putting the laser source and photonic chip on the bench.

IV. Conclusion

A code Python has been developed to manipulate the characterization bench and the tools necessary to characterize a divergent and elliptic laser beam. Even if the bench could not yet be tested on a laser source and OPA, it showed it's edge coupling capacity with another laser source and photonic chip combination. By using the tool developed for the bench, the different approximations allow for a more precise approximation of the laser position and orientation, as well as it's position in respect to the edge coupler. The Fourier IQA as shown itself to be the most reliable of the 2 IQA methods and the setting for laser source orientation are a Gaussian filter of size 5 and a step of 0.1° for the angles searched between -21 and 21°.

V. Discussion

This paper layout the basis of how to do a edge coupling characterization bench for a laser source elliptic and divergent. But to fully automatized the bench, a detection method to automatically find the laser source and edge coupler on the image is still needed. The entropy IQA probably failed to find the best focus due to the strange nature of the image used, even if the time taken for each IQA is consistent with the algorithm complexity from the articles. For the laser source position found after one cycle, it is important to note that the laser source diameter is 100 µm, making the error in measurement still inside the optical fiber. Furthermore the distance found on the Z_W-axis correspond more likely to an error in measurement more than an actual distance. The distance found with this method on the Z_W-axis should be compared with another method to see if the results is really reliable. The force detection method should be improved as well, or another piezo-motor used. Ideally 1.96 mN should be enough to detect a contact, as it is the force used to manipulate laser source and photonic chip in laboratory. A GIU should be developed as well to make the bench more easy to use.

Acknowledgment

The author would like to thank the supervisors: Sylvain Guerber and Sebastian Repeztky for their help, interest shown and advice during the assembly of the optical characterization bench. The author would like to thank as well all the members of the LIPS of the CEA-Leti, especially Olivier Castany and Phillipe Gross who shared a lot of their time to help with their experience and helpful discussion.

References

[1] B. S.-R. A. R. G. P. H. C. A. D. M. Ching-Pai Hsu, Boda Li and V. Donzella, "A review and perspective on optical phased array for automotive lidar," *IEEE*, 2019.

[2] I. C. P. G. K. A.-C. J. F.-T. Sylvain Guerber, Daivid Fowler and B. Szelag, "Development, calibration and characterization of silicon photonics based optical phased arrays," *SPIE - The International Society for Optical Engineering*, vol. XXIII, 2021.

[3] R. Marchetti, C. Lacava, L. Carroll, K. Gradkowski, and P. Minzioni, "Coupling strategies for silicon photonics integrated chips," *Photon. Res.*, vol. 7, no. 2, pp. 201–239, Feb 2019. [Online]. Available: http://www.osapublishing.org/prj/abstract.cfm?URI=prj-7-2-201

[4] G. T. Reed and A. P. Knights, *Silicon photonics an introduction*. Wiley, 2007.

[5] K. H. John X.J.Zhang, *Molecular Sensors and Nanodevices*. Wiley, 2014.

[6] L. J. Domagoj Jakobovic, "The forward and inverse kinematics problems for stewart parallel mechanisms," 2002.

[7] E. M. B. Mauricio Becerra-Vargas, "Application of h_∞ theory to a 6 dof flight simulator motion base," 2012.

[8] V. H. G. J. Rajevenceltha, "A novel approach for image focus measure," *IEEE photonics journal*, 2020.

[9] K. De and V. Masilamani, "Image sharpness measure for blurred images in frequency, domain," 2013.

[10] J. CANNY, "A computational approach to edge detection," 1986.

[11] R. O. Duda and P. E. Hart, "Use of the hough transformation to detect lines and curves in pictures," *communication of ACM*, vol. 15, 1972.

[12] S. H. H. Sanidhya S. Naikwad, Ruben Vandervelden, "Self-sensing contact detection in piezo-stepper actuator," *IEEE*, 2016.

Germain Favier is following a double master between MCI Innsbruck/Austria in Mechatronics and Smart Technologies and ESILV Paris/France in is a student at MCI Innsbruck/Austria. Among others he is responsible for the activities in control engineering and fluid mechanics in which he is regularly publishing.

Measurement System for a Laboratory Granular Chute to Determine the Flow Behaviour of Gravity-Driven Mass Movements

Simon Felbermayr[†], Anselm Köhler[‡*], Engelbert Gleirscher[‡*], and Bernhard Hollaus[†*]

***Abstract*—Measurements on gravity-driven mass movements in nature is a complex task, moreover, real-scale test sites and gathered data are rare. Considering mass movements as granular flows, scaled laboratory experiments are often used to investigate parameters of the dynamics and the flow behaviour. Therefore the goal of the work is to generate automated 2D velocity field and flow depth information in a scaled mass chute.**

Two optical measurement techniques are investigated and compared with previous done analogue measurement techniques. Optical Flow (OF) is used to infer the velocity field, which is compared to Optical Speed Measurement (OSM) sensors using UV-diode pairs. The flow depth is estimated using Structured Light Illumination (SLI) technique along a laser line and reference is taken from distance measuring point lasers. A high-resolution 4K camera with a framerate of $160\,\mathrm{fps}$ is installed perpendicular to the chute surface. Custom codes are developed for controlling, data acquisition and post-processing. Further components like customizable LED lighting, an electronic tilt mechanism to change the chute angle and an automated granulate inlet are installed. The later reduces the oscillations of the chute to a minimum and ensures the synchronisation of the measurement system.

The OF algorithm enables to track velocities up to $31\,\mathrm{m\,s^{-1}}$. The evaluation of the SLI shows that flow depths in the order of 5 to $10\,\mathrm{mm}$ can be resolved. However, a minimum of $100\,\mathrm{mW}$ laser power is needed to retrieve continuous depth readings of the dark granulate material.

The developed measurement system is able to determine the spatio-temporal flow behaviour (velocity field, depth information) of granular flow on a laboratory scale. To compare scaled experiments with real mass movement event, the dimensionless Froude number is an indicator. With the performed experiments, a Froude number of 7.09 can be obtained that is in the range known from cohesionless cold-dry dense flow regimes to even fluidized flows within the denser parts of powder snow avalanches.

***Index Terms*—Avalanche Chute, Granular Flow Dynamics, Optical Flow, Structured Light Illumination, Particle tracking.**

Correspondence: felbermayr.simon@outlook.com
[†] Department of Mechatronics, Management Center Innsbruck (MCI), Innsbruck, Austria
[‡] Department of Natural hazards, Austrian research institute for forest (BFW), Innsbruck, Austria
[*] supervisors
Submitted: July 27, 2022

I. Introduction

To determine the flow behaviour of gravitational driven natural hazards like snow avalanches and soil flows, real scale field measurements, simulations and laboratory scaled experiments are carried out [1]. In such a scaled environment, different granular material must be used to substitute for the snow in an real avalanche, as the flow rheology can not be matched with real snow in such small scales [2]. For comparing laboratory scale and real scale mass-movements, dimensionless scale quantities have to be fulfilled [3]. The Froude number as the ratio of the flow inertia to the external gravitational field is of high importance. When the Froude number is similar, scaled granular flows can be used to investigate the interaction with obstacles like nets or dams and the results can indicate the effectiveness of such retarding structures.

Throughout this work a laboratory measurement chute is developed that is named AvaChute. The already existing chute designs and tests are described in [4] and [5]. Measurement equipment like Optical Speed Measurement (OSM) sensors, point lasers and a camera are used already in the previous chute designs. The original camera with 24 fps was not able to fully track the flow, but the front velocity could be extracted from the frames manually. The OSMs and the point lasers were evaluated with a LabView software, giving local flow velocity and flow depth information at single points.

The aim of the new chute setup is to get a automated spatial 2D velocity field instead of only the front velocity by means of Optical Flow (OF) analysis, as well as estimate flow depth information over larger areas of the chute with the method of Structured Light Illumination (SLI). For this purpose a high-resolution 4K camera with 160 fps is used. An optical flow algorithm evaluates consecutive frames and infers the velocity field. Further for the flow depth line lasers are used to apply in the post-processing SLI technique. From the initial chute the OSM sensors and point lasers are kept in the current setup as reference, but a new post-processing program is applied. The overall goal of the measurement chute setup is to verify a Froude number ranging between 3 and 8 as this range includes cohesion-less cold avalanches [6].

II. Methods

A. AvaChute Setup

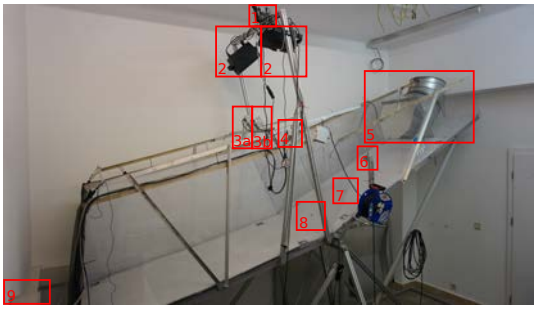
The redesigned chute is shown in Figure 1a, wherein the used components are marked and listed in Table I. All dimensions of the $4.5\,\mathrm{m}$ long and $1\,\mathrm{m}$ wide chute are shown in Figure 1b, also indicating both angles α and β of the lower and upper chute segments, respectively. All electrical components are routed to a central control box mounted to the wall. This box contains the distribution of the DAQ, the relays and the voltage supplies. This box is further connected to the DAQ card, and some USB ports of the control PC. A custom Python program is made to control the sensors and data acquisition as well as the post processing.

TABLE I: AvaChute components and sensors

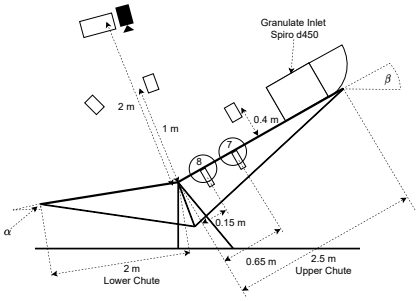
Nr	Components	Type
1	Camera	Z-Cam E2
1	Camera lens	Panasonic Lumix G Vario 7-14 mm
2	LED Strobe light	Stairville xBrick Quad $16 \cdot 8\,\mathrm{W}$ RGBW
3a	Point Laser 2	Balluff BOD 63M-LA04-S115
3b	Point Laser 1	Balluff BOD 63M-LA04-S115
4	Line Laser Pos 1	LLMiHybrid 12mm 20 mW / 100 mW
5	Granulate inlet	Spiral duct d450 mm 45° / Flap: checker plate with electrical door lock BEFO 11211MB
6	Line Laser Pos 2	LLMiHybrid 12mm 20 mW
7	OSM 1	Developed by [7]
8	OSM 2	Developed by [7]
9	Control Box	HENSEL 6547

The analogue to digital conversion of the OSMs and point laser is performed with the NI-DAQ 6024e, which is placed in a 32-bit PC with a Windows operating system. The NI-DAQ 6024e is placed inside the PC and a terminal block NI-CB-68LPR is used and situated in the central control box that connects all peripherals. All sensors are sampled with 20 kHz and are measured as Referenced-Single-Ended (RSE). The sampled data is stored into an HDF5-file for further processing.

All measurement systems can be divided into two main groups: the flow depth and flow velocity. Furthermore, each group divides into point measurements from the previous setup and into the spatial measurements that are developed throughout this work. These measurement systems and their working principles are discussed in the following sections.



(a) Picture of the experimental setup with components 1-9, that are listed in Table I.



(b) Schematic drawing of the chute with the relevant dimensions.

Fig. 1: Overview of the new AvaChute, that is basically a short version of the original chute setup. Not only renewed is the measurement equipment, but also the release and tilt mechanism are automatic.

B. Chute Lighting

Two Stairville xBrick Quad modules with each $16 \cdot 8\,\mathrm{W}$ RGBW LEDs are installed above the chute on both side to minimise shadows. The lighting of the chute can be can be controlled by software (QLight-Controller+). For the experiments several lighting schemes with the colours blue, green and RGBW, and always on and pulsed light state are tried to find the optimum for all methods. In the case of the pulsed lighting scheme, OF is applied during the bright phase whereas for the SLI the dark phase are necessary for extraction of the line laser reflections. Since the ZCam uses a rolling shutter some frames are not usable when the shutter is partly closed or opened during a light change. These frames are therefore excluded from the flow velocity calculation as well as for the flow depth estimation. If the light is pulsed for the experiments, the bright phase T_{L-on} is set to $210\,\mathrm{ms}$ and the dark phase T_{L-off} is set to $40\,\mathrm{ms}$ equating to 4 cycles a second.

C. Point Laser Measurements

Reference flow depth measurements are taken with the Balluff BOD 63M-LA04-S115 point laser sensors. These sensors work by the principle Time-of-Flight (TOF). The calibration of the sensors are done by taking reference objects on the chute and back calculate the conversion factor which is then applied to the analogue signal of the laser sensor. This factor is found to be $0.02\,\mathrm{V\,cm^{-1}}$.

D. Optical Speed Measurement Sensors

The Optical Speed Measurement (OSM) sensors were developed by [7]. The measuring principle is that emitted Ultraviolet (UV) light is reflected by a passing object and further sensed by a photo diode. Two pairs of UV-LED and photo diode are needed to measure the velocity and placed with a separation of $11\,\mathrm{mm}$ inside one OSM. The analogue signals of both diodes are then cross correlated to find the associated velocity [7]. Previous work with the granulate chute already showed, that the diodes are not well suited for dry granulate since they were mainly tailored for wet snow reflectance properties, thus, the cross-correlation mostly returns only a reliable value for the passing of the flow front [4].

E. Camera and Image Calibration

The camera frames have two distortions. The first distortion is caused by the camera and the lens and can be described by the intrinsic and extrinsic distortion

matrix following the pinhole model [8]. The second distortion is due to the geometrical arrangement of the camera in respect to the AvaChute surface. Note, the chute contains two elements with the specific angle α and β as shown in Figure 1, thus, for each element a different correction is applied.
In this work the intrinsic and extrinsic matrix are found by searching chessboard corners using functions from the `openCV` library. Once this distortions are corrected, there is still the need to address the perspective and scaling of each frame. ArUcos marker symbols are glued to all four edges of each chute element. Each of the ArUcos markers has an individual ID giving the advantage to know which one is placed at which corner on the chute. With this information the frames perspective are corrected. Scaling the image, e.g. matching the pixel size to a real world length scale, is done by searching again the chessboard corners but now in the undistorted frames that contain the correct perspective. Therefore the mean distance between the chessboard corners for the width and length of the image is used to find the scaling factor for both dimensions.

F. Optical Flow

For the Optical Flow (OF) calculation the Gunnar–Farnebäck algorithm is used [9]. Basically, the OF method is better suited for continuous moving material without discrete tracer particles compared to particle image velocimetry [10]. In Python the openCV function `calcOpticalFlowFarneback` is applied to two consecutive frames. The quality of the optical flow calculation depends mainly on the input parameters beside the image quality. In this work the main focus is set to six input parameters, namely:

- `pyr_scale`: Image scale to build pyramids
- `levels`: Number of pyramid layers
- `winsize`: Average window size
- `iterations`: Iterations performed at each level
- `poly_n`: Pixel neighbourhood
- `poly_sigma`: Standard deviation of Gaussian filter

The mentioned declaration and description of the input parameters are taken from [11] and will be described throughout this section.
In the AvaChute a maximum velocity of $5\,\mathrm{m\,s^{-1}}$ is assumed, therefore the granulate front jumps over many pixels in consecutive video frames. Algorithms like the one from [9] are having troubles with large pixel offsets, therefore multi scale displacement estimation [9] is used. [11] provides a technique named Image Pyramid to observe down-scaled copies of the input image to calculate the flow. With such down-sampled images the flow can be determined easier for big movements of objects. With the outcome of the down-sampled image the algorithm is stepping one level higher and has a initial flow velocity to track the movement in the true image resolution [12]. Here the scaling factor `pyr_scale` and the levels of scaling `levels` are taken as a input of the function `calcOpticalFlowFarneback`.
In addition, the stability of the result can be influenced by setting the parameter of the window size `winsize`. On the one side with a growing window size increases the stability but on the other side the output is more blurred over the whole velocity field. The algorithm implemented in [11] also takes `iterations` as an input parameter, which defines the maximum iterations performed in each pyramid level to reach the convergence. By increasing the neighbourhood size `poly_n` the velocity field gets more smoothed but in addition the algorithm will be more stable and robust. This parameter is used to find a polynomial expansion in the pixels for the Farnebäck algorithm [9][11]. The last parameter `poly_sigma` defines the standard deviation for the Gaussian filter which is applied over the neighbourhood to reduce noise in the images.

TABLE II: Input parameter of optical flow algorithm

Parameter	Value	Parameter	Value
pyr_scale	0.7	poly_n	7
levels	10	poly_sigma	1
winsize	16	flags	0
iterations	3		

Combinations of these parameters were tried and their results are discussed in the accompanying master thesis [13]. The found optimal parameter set is listed in Table II, and are used for all velocity calculation mentioned in subsection III-A.

G. Structured Light Illumination

The Structured Light Illumination (SLI) principle is shown in Figure 2. The red line laser is placed with a defined angle ϕ sideways to the chute. Whenever a object passes through the projected line, the line seems diverted when seen from a different angle. The camera is perpendicular to the chute and shows the diverted laser line as an offset Δs from the top view as sketched in Figure 2. The depth d_i

$$d_i = \frac{\Delta s}{\tan \phi} \tag{1}$$

of the moving object is evaluated from Δs with a trigonometric relation for each pixel i where the laser beam is identified [14].

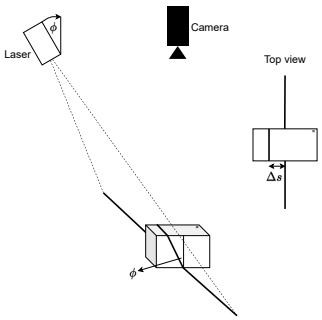


Fig. 2: Principle of SLI

The development process in [13] shows that HSV - thresholding method gives the best results when creating a binary mask to extract the red laser of an image. The colour space HSV stands for hue, saturation and value and is chosen because all red-like colours can be more easily identified by thresholding compared to using for example RGB - colourspace where a mixture of all RGB create the red-like colour spectrum of different intensity.

III. Results

Throughout this section, the main results of the velocity estimation and the depth calculation are presented, but only a selection of all performed measurements are shown. The granulate material used for all the experiments is Neopor particles, that are graphite coated plastic spheres with a diameter of 1.3-$6.5\,\mathrm{mm}$, in a mixture with glass spheres of diameter 1.5-$2\,\mathrm{mm}$ [4]. Roughly a volume of approximately $0.007\,\mathrm{m}^3$ is used for the experiments.

A. Velocity Profile

The input parameters for the optical flow algorithm that gives the best results throughout the whole flow evaluation are listed in Table II and are used here. The result of the OF algorithm is shown in Figure 3, that gives one snapshot of the 2D velocity field of the flow at the specific timestamp of $550\,\mathrm{ms}$ after the granules are released. The image is calibrated as mentioned in subsection II-E, and here only shows the upper chute segment. Panel a) is showing the first input images with the grayscale intensity i_{gray}, panel b) is the absolute velocity v_{abs} field that originates from the resolved horizontal velocity v_x and vertical velocity v_y field as shown in c) and d), respectively. The granulate is released from the granulate inlet at the right side in the middle and the granulate is flowing to the left, therefore the horizontal velocity is negative.

The red dot plotted at the grayscale image in Figure 3a) indicates the point where the velocity evolution over time of the passing granular flow is evaluated. The absolute velocity v_{abs}, the velocity in x direction v_x and the velocity in y-direction v_y over time t are shown in Figure 4 together with the reference velocity from the OSM v_{OSM}. For the first granules passing the OSM, the front velocity is calculated from the analogue signal. Note, that later in the flow any correlation is not well defined for the OSM and no other reference velocity values can be estimated. This velocity is plotted at the time $0.56\,\mathrm{s}$, whereas the front in the OF calculation reaches the image point at $0.69\,\mathrm{s}$.

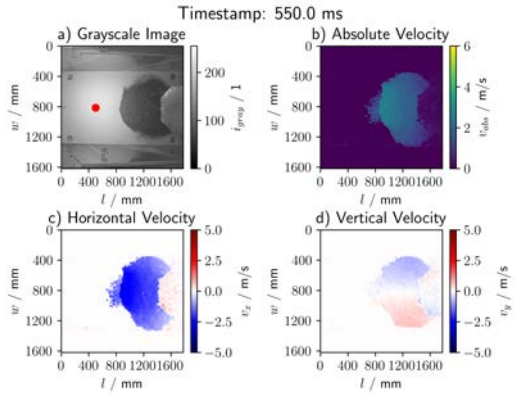


Fig. 3: Velocity field of the granular media resolved with the optical flow method at timestamp $550\,\mathrm{ms}$. Panel a) shows the first of both input images. Panel b) shows the absolute velocity v_{abs} calculated from the horizontal velocity v_x in Panel c) and vertical velocity v_y in Panel d)

This time gap is on one hand caused by the synchronisation error of the DAQ and the camera. Out of the log files in python it shows that the synchronisation error of the camera and DAQ is about $50\,\mathrm{ms}$. On the other hand slightly spatial difference in the optical flow algorithm coordinates and the OSM analogue response time adds further up to the time shift.

The velocity of the OSM from the cross-correlation of the analogue signal and the front velocity are in the same order of magnitude of 2–$2.5\,\mathrm{m\,s^{-1}}$. A small difference is always expected due to the difference between the surface and slip velocity of a granular flow, e.g. between the OF-derived surface velocity from the camera above to the slip velocity measured with the OSM from below.

Obviously, the velocity signal behind $1.5\,\mathrm{s}$ appears to contain noise. But the flow is very thin towards the tail and sometimes only single granules slide down the chute that can lead to different velocities. These fluctuating velocities can be realistic for single particles that undergo individual collisions, but also the optical flow algorithm is not well suited for several single particles inside a window size `winsize` that is chosen optimal for the dense flow.

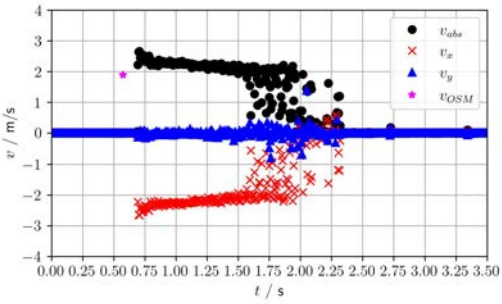


Fig. 4: Velocity evolution at point [$l = 500$, $w = 815$] over time. Shown is the absolute velocity v_{abs}, velocity in x direction v_x and the velocity in y direction v_y derived from the OF algorithm. Further velocity v_{OSM} at location of the OSM 1 sensor is plotted as reference.

B. Depth measurements

Two line lasers with $20\,\mathrm{mW}$ and $100\,\mathrm{mW}$ output power and a aperture-angle of $120°$ are investigated. They are mounted on the chute in a way that the lines are projected parallel to each other on the chute surface. Needless to say, that the thin red laser line is only visible when the strong light for the OF analysis is turned off. Four methods were investigated to convert the camera image into a binary mask containing pixels with the laser light. This comparison is not shown here, but is given in the accompanying master thesis [13]. Most robust is the laser line extraction using HSV-thresholds, which is used for the following analysis.

The laser line is a couple of pixels wide on the glossy surface texture of the white chute material, therefore several depth values d_i are resolved for each pixel at the same chute length l. We defined the spread of the line orthogonal to the orientation of the original line,

e.g. spread of the binary mask in cross-slope direction of the chute, as the standard deviation std over all d_i. The mean depth d and the standard deviation std over chute length l are shown for the chute surface and the granular flow in Figure 5.

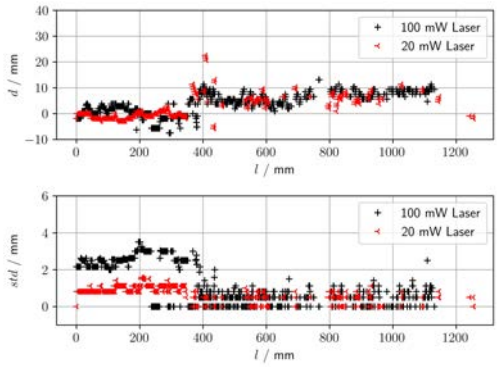


Fig. 5: Depth measurement with no lightning and granular material at upper chute with $20\,\mathrm{mW}$ and $100\,\mathrm{mW}$ laser. Granular media is flowing from right to left and has reached approximately $390\,\mathrm{mm}$.

The granular front at $400\,\mathrm{mm}$ divides the data into purely the chute surface for small and granulate cover for larger length l values. For the chute surface, the average depth is $0\,\mathrm{mm}$ with an uncertainty of about $\pm 3\,\mathrm{mm}$. The $100\,\mathrm{mW}$ laser causes more reflections and the binary mask is wider giving a std of up to $3\,\mathrm{mm}$, thus an average width of the laser line of about $6\,\mathrm{mm}$ or pixels.
On the black matte granulate, the spread of the laser line diminishes for the $100\,\mathrm{mW}$ laser and is similar to the $20\,\mathrm{mW}$ laser line as the std for both is comparable. However, there are a couple of sections along the granulate where the $20\,\mathrm{mW}$ laser line can only be detected sporadically causing data gaps. Such gaps are for example at length l of 450 to $530\,\mathrm{mm}$, 610 to $790\,\mathrm{mm}$ and above $970\,\mathrm{mm}$. The stronger $100\,\mathrm{mW}$ laser gives a more continuous set of depth measurements along the granulate flow and defines the minimal required laser line strength for this particular SLI application. The depth resolution is about $5\,\mathrm{mm}$ which is basically the granule size.
As reference, the Balluff point laser gives depth information at one specific coordinate in the chute throughout the measurement. Comparison of the Balluff point laser and the $100\,\mathrm{mW}$ line laser depth d over time t is shown in Figure 6. The point laser's raw analogue signal is depicted as the gray area, from which a moving average with a window size of 500 samples is calculated to smooth out the noise in the data and results in the blue dashed line. The line laser is evaluated at the same length but at a different width in the image plane since the line and point laser can not be focused to the same location. This means the line laser and point laser measure slightly different parts of the flow with a gap of approximately $10\,\mathrm{cm}$, but at least the order of magnitude in flow depth should be comparable.

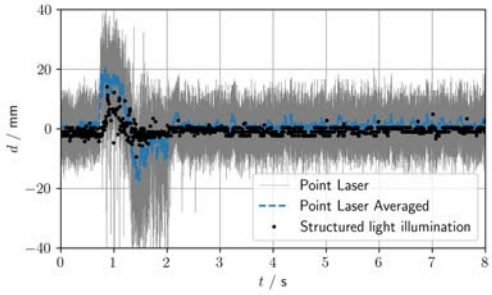


Fig. 6: Depth information of Balluff laser sensor and $100\,\mathrm{mW}$ line laser. The two lasers are evaluated for one specific coordinate in the chute, but with a gap of $10\,\mathrm{cm}$ in width to each other.

The granular media reaches the point laser at a time of $0.78\,\mathrm{s}$. Here the point laser increases to the frontal depth of the granulate. After that a negative depth is reached, this is caused by a paper background on the chute that is not totally flat and the granulate flow may depress the paper. The line laser response shows that the granular media front reaches the coordinate at time $0.84\,\mathrm{s}$. Also here some negative measurements

are visible after the front at $1.4\,\mathrm{s}$.
Comparing the depth results of the point laser with the depth derived from the line laser, it shows that they differ a little from each other. The line laser detects a depth about $10\,\mathrm{mm}$ for the front at approximately $0.87\,\mathrm{s}$, whereas the point laser shows a depth of about $14\,\mathrm{mm}$. This difference is likely be caused since the sensors are measuring a different front due to the sensors offset in chute width, thus, the flow depth or mass distribution must not be the same. Similar argument explains the time difference when the granulate front reaches the sensing location. However, a second reason is the synchronisation error which also has to be taken into account by means of post-processing.

IV. Discussion & Conclusion

The main motivation to carry out small scale laboratory chute experiments are to gain more knowledge about real world mass movement events and study the interaction with for example mitigation structures. Therefore the dimensionless Froude number is a appropriate approach to check the scaling behaviour of these laboratory flows with the real-scale equivalents.

Using the above derived values for the mean flow depth H and the mean flow velocity v_x, we find a Froude number

$$F = \frac{v_x}{\sqrt{g \cdot H}} = \frac{2.22\,\mathrm{m\,s^{-1}}}{\sqrt{9.81\,\mathrm{m\,s^{-2}} \cdot 0.01\,\mathrm{m}}} = 7.09 \tag{2}$$

at the upper end of the desired range for the Froude numbers between 3 and 8, that correspond to a rather turbolent and fluidized flow. Such a Froude number indicates that cohesion-less cold snow avalanche or even more agitated flows inside of a powder snow avalanche can be replicated in this scaled chute. The adjustablility of the chute, e.g. angle, chute surface, granulate material and volume, allows to adjust the Froude number in both positive or negative directions.
Note, that the mean flow depth H is taken from Figure 5 at the length $400\,\mathrm{mm}$ and the mean velocity v_x from Figure 4 at the specific timestamp of $693.75\,\mathrm{ms}$, thus, from two different experiments. As a first approximation, this should be giving valid results, since flow repeatability can be assumed throughout the experiments, e.g. two experiments with the same chute angle and release volume should give the same flow field. This step is necessary, since the velocity estimation with OF and the depth estimation with SLI have basically opposite lighting requirements.
The work on the new optically-based sensing equipment for the chute setup reveal that a very specific lighting is crucial for successfully processing the video data for the two methods. Basically, the main result here is that both methods can not work simultaneously together, and certain drawback have to be accepted. OF needs very bright light to capture the fast moving granules, SLI needs basically no light to best see the faint reflections of the red laser. Therefore, a tradeoff is found with a pulsed lighting scheme even though the measurement of SLI and OF can only be done alternately. However, the pulsed light completely saturates the analogue signal of the diodes and turn the OSM signal unusable.
Nevertheless, the two herein developed sensing methods that base on a high-resolution 4K camera with a sufficient high framerate of $160\,\mathrm{fps}$ improves the AvaChute significantly. Now, either a complete spatial 2D velocity field or the flow depth along a laser line can be resolved with $160\,\mathrm{Hz}$. Or if using the pulsed lighting setup, one alternates between measuring flow velocity and flow depth four times a second.
There is still room for improvements for all developed methods and continuing work will enhance the data quality. The time synchronization need to get investigated in more detail, however, the analogue sensors (Balluff and OSM) are already less important compared to only camera data. The flow depth measurement with the line laser should improve when either increasing the laser power further, or by testing differently reflecting granulate material.
This work presents the successful implementation of two spatio-temporal methods to assess the granular flow behaviour in a scaled laboratory environment.

The new measuring system of the AvaChute is a powerful and reliable tool to model mass movements and compare the results with the real world and with numerical simulations. Future plans are to include 3D terrain contours into the chute to work on terrain dependent questions, to use the lower angled chute segment for studies on deposition and entrainment, and to add a conveyor belt or similar granulate feeder to enable steady state flows and virtually extending the $4.5\,\mathrm{m}$ long chute, to make the AvaChute truly unique.

ACKNOWLEDGMENT

This project has been financed by the Unit of Snow and Avalanches (BFW) Innsbruck.

REFERENCES

[1] F. Tiefenbacher and M. Kern, "Experimental devices to determine snow avalanche basal friction and velocity profiles," *Cold Regions Science and Technology*, vol. 38, no. 1, pp. 17–30, jan 2004. [Online]. Available: https://doi.org/10.1016%2Fs0165-232x%2803%2900060-0

[2] B. Turnbull and J. N. McElwaine, "Experiments on the non-boussinesq flow of self-igniting suspension currents on a steep open slope," *Journal of Geophysical Research*, vol. 113, no. F1, jan 2008. [Online]. Available: https://doi.org/10.1029%2F2007jf000753

[3] J. Koegl, A. Graf, L. Rammer, K. Kleemayr, M. Kern, P. Gauer, G. Kapeller, and M. Aufleger, "Scaled laboratory experiments on the evolution of fluidised avalanches," 09 2009, pp. 477–481.

[4] E. Gleirscher, "Gutachten Mühlauer Klamm Teil 2 - Modellversuche zur Bestimmung des optimalen Standortes und der Bauwerkshöhe des 2. Lawinenbremsbauwerkes in der Mühlauer Klamm," Department of Natural Hazards, Hofburg Rennweg 1, A6020 Innsbruck, Tech. Rep., May 2012.

[5] E. Gleirscher and J.-T. Fischer, "Retarding avalanches in motion with net structures," *Cold Regions Science and Technology*, vol. 97, pp. 159–169, Jan. 2014. [Online]. Available: https://doi.org/10.1016/j.coldregions.2013.08.007

[6] A. Köhler, J. N. McElwaine, and B. Sovilla, "GEODAR data and the flow regimes of snow avalanches," *Journal of Geophysical Research: Earth Surface*, vol. 123, no. 6, pp. 1272–1294, Jun. 2018. [Online]. Available: https://doi.org/10.1002/2017jf004375

[7] Gerrit Günther, "Entwicklung eines optischen Sensors zur Messung von Fließgescchwindigkeiten in Slush Flows: Praktikumsbericht des 2. praktischen Studiensemesters," Eidg. Institut für Schnee- und Lawinenforschung SLF, Tech. Rep., 31.3.2006.

[8] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 11, pp. 1330–1334, 2000. [Online]. Available: https://doi.org/10.1109/34.888718

[9] G. Farnebäck, "Two-frame motion estimation based on polynomial expansion," in *Image Analysis*, J. Bigun and T. Gustavsson, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2003, pp. 363–370.

[10] B. K. Horn and B. G. Schunck, "Determining optical flow," USA, Tech. Rep., 1980.

[11] G. Bradski, "The OpenCV Library," *Dr. Dobb's Journal of Software Tools*, 2000. [Online]. Available: https://opencv.org/

[12] J. Bai and L. Huang, "Research on LK optical flow algorithm with gaussian pyramid model based on OpenCV for single target tracking," *IOP Conference Series: Materials Science and Engineering*, vol. 435, p. 012052, Nov. 2018. [Online]. Available: https://doi.org/10.1088/1757-899x/435/1/012052

[13] S. Felbermayr, "Measurement system to determine the flow behavior of natural hazards like granular media with a laboratory scale chute," Master's thesis, Management Center Innsbruck, Maximilianstraße 2, July 2022.

[14] R. A. Wagh, M. S. Panse, and H. Apte, "Calibration method for height measurement of object using laser triangulation," *International Journal of technology Enhancements and emerging Engineering Research*, vol. 3, no. 04, p. 27–31, 2015.

Simon Felbermayr is a master student at the Department Mechatronics and Smart Technologies with the specialisation to electrical engineering at the Management Center Innsbruck. He succeeded his bachelor degree with developments for a new snow avalanche radar system.

Design of a Neural Network to Determine the Number of Turns and Jumps of Mountain bike Descent

Maximilian Karl Freiheit, and Bernhard Hollaus (supervisor)

***Abstract*—The following paper describes the neural network approach with a CNN-LSTM architecture which is adapted for the estimation of the number of turns and air time on a mountainbike descent, determined out of mobile phone sensor data. Due to the possibility of the data acquisition with mobile phones, personal tracking of a bicycle activity can be brought to a next level, which is why this thesis uses data generated with mobile phones.**

Currently, training data is most widely based on GPS data. Recorded GPS based track data can only deliver data on duration, distance, velocity and elevation. However, this does not represent the whole difficulty of a track since the number of turns, as well as the airtime, are not measured. With the use of accelerator, magnetometer and gyroscope data, generated by a mobile phone, a neural network can transform the sensor information of a recorded session. This makes it possible to both evaluate and compare different tracks at a whole new level.

To enable this improvement, an algorithm is determined. This is done by means of research to find the algorithm best suited to determine the number of turns and airtime from mobile phone sensor data. Next, predefined mountainbike tracks are used to acquire the track data with the rider carrying a mobile phone with him, which records sensor data. This phone can be attached to the mountain bike stem or carried in various pockets. Further, the data that is recorded on this route is accurately labelled, in order to learn, train and test the algorithm. This is done manually with a self written matlab application and video material of the ride. The CNN-LSTM based neural network is then optimized in regards to its accuracy with the batch size, the window length of the sliding window approach and Savitzky-Golay filter is used. This filter did not improve the accuracy of the network. Finally, the prediction of the trained algorithm is evaluated. The final output of the neural network is post processed and compared to the labelled input. Further, results are discussed and an outlook and suggestions for future improvements are given.

The obtained data and tested neural network show that it is possible to determine a tendency for specific features, like a turn or a jump out of the data collected. However, with the for this thesis acquired amount of data, the algorithm could not be trained efficiently enough to lead to a clear classification of the data. The result of seeing the capability of a neural network spurs the motivation to collect and label more data and train the algorithm further. This would make the algorithm more precise and enable a deeper comparability of different mountainbike tracks in terms of their difficulty.

In conclusion, it can be said that with the help of a neural network, the next level of training evaluation can be obtained. But to gain a real precise prediction out of a trained network, a larger amount of data is needed for the training. Apart from the lack of data, it can be seen, that the approach described in this master thesis works in the real application of a mountainbike ride, which also enables the rider to store his mobile phone independently.

***Index Terms*—neural network, sports tracking, track estimation, personal tracking, downhill parameter estimation, mobile phone data processing, LSTM, RNN, CNN, multi label classification, activity classification**

I. Introduction

AT the beginning, the motives for a curve and jump detector by means of mobile phone data are described. Then, the current situation is shown and finally, the procedure and methodology used is explained. In general, the usage of a mobile phone for the data acquisition is applied to make it possible to easy generate data sets. Further, this enables the finally trained algorithm to be used by every mountain biking mobile phone owner.

II. State of the art mountain bike tracking

In first instance, it is to mention that the kind of tracking is divided into two groups: The tracking with special devices and the tracking that can be established with a mobile phone application. Hardware tracking can be conducted with a variety of devices. The mainly used ones are GPS bike computers and wearables such as smart or sport watches. The manufacturer of navigation hardware, Garmin, already gives a solution to the need of comparability of tracks with its special devices. For this, Garmin implemented the MTB Dynamics Parameters in some of its wearables and bike computers. Besides tracking through specific hardware, there is also the possibility of software solutions for mobile phones. These well known cycling tracking applications for mobile phones, e.g. strava, runtastic mountain bike, trailforks and komoot do not provide any function for evaluation apart from the main parameters as previously defined. One application, the author found during his research provides deeper analyses of the ride. This application is named MTBHangtime. The same called app delivers the airtime and the inclination of a mountain bike while riding [1]. However, it does not provide the count of turns. At the same time, the mobile phone has to be mounted to the mountain bike which restricts the user cases as it does demand the rider to again purchase specific hardware, even though only used to mount the phone .

III. Methodology

The chapter Methodology explains all the relevant steps conducted to determine an algorithm that can extract maneuvers out of mobile phone sensor data of a mountainbike ride. First, the arrangements for the acquisition of the relevant data is given. Next, the exact labelling procedure is detailed. Completing this section, the relevant steps of preprocessing for a neural network are described. The determined Neural Network and its structure is shown in Figure **??**. The at the moment common mobile phones are mostly smart phones, which have three built-in sensors [2]. These sensors include a sensor to measure the acceleration in three axis, a magnetometer to measure the orientation of the mobile phone according to the location on the earth and a gyroscope to measure the angular velocity. This sensor data is recorded during several experiment rides and saved in a comma separated values file, short - .csv-file. All the sensor data is mainly recorded with a frequency of 100 Hz due to the fact that this is the maximum possible recording rate of the author's mobile phone. Artificial Neural Networks have the ability to analyse big data sets and find specific events within them [3]. Hence, these Artificial Neural Networks are able to classify completely unknown data sets into predefined classes. Due to this attribute, a Neural Network is chosen for the task of classification of mobile phone sensor data into different riding maneuvers.

A. Data acquisition

As said, the data acquisition is done with a general mobile phone. For the propose of the experiment, a mobile phone is positioned on the stem of the bike or in the riders pocket. The data set collected and used for the algorithm training encompasses 781.687 samples with 12 features of pure descending track data. The gathered sensor data is recorded using a frequency of 100 Hz. Using the highest possible sampling frequency is of great advantage, as it also generates the highest amount of data possible. Generation of a higher amount of data also

generates a higher amount of training data, which can be used for the neural network. Therefore, a more accurate algorithm can be developed. For the recording, as mentioned, the hardware of a mobile phone is used. Using the open source application ”phyphox”, developed by the RTWH Aachen, sensor data is recorded during the experiment. The application is set to record with a sampling rate of $100Hz$. This is due to the maximum capability mobile phone mostly used within this experiment being an iPhone 11 Pro. To ensure a uniform data set of all recordings and to add simplicity, all acquisition is done with $100Hz$. Whilst recording, the mobile phone is stored in different positions during the various rides. This is done in order to generate an overall more universal data set.

Fig. 1. Mobile phone mounted to the stem

Figure 1 shows the acquisition set up on the stem of a mountain bike. This is one of three used mounts.

B. Video processing and Data labelling

As machine learning is driven by the amount of data acquired as well as the quality of the obtained data, the larger the data set and the more precise the labelling, the more efficient the algorithm can perform. When labelling the acquired data, it is crucial for each sample to be given a label for the case of the used supervised learning method. This means that every recorded data point, which is represented by one row with up to 12 features, has to be assigned to one of the investigated maneuvers. At the same time, events may not be mislabeled, as this would lead to wrongful weighting in the training of the neural network and decreases its accuracy in the prediction of unseen data. It is the objective of this thesis to develop an algorithm which classifies the obtained mobile phone data into curves and jumps and counts all amounts of these maneuvers. Therefore, the ride is recorded with an action camera. The video footage enables the data points to later be assigned its correct maneuver classification manually. The following four classes have been decided to be labelled on the data: neutral, jump, left turn and right turn. To assign a label to every data point, the footage of the action camera, in its initial state, has to be synchronised to the sensor data. This is done by cutting of the video to the start of the sensor data acquisition, which is visibly in the video through filming the click on the mobile phone. Additionally, the video end is cut off when the rider starts to rest. The validation application will append the given labels to the sensor data. It must be mentioned, that no personal information of the rider is stored. This ensures ethical correctness for the experiment.

Fig. 2. Screenshot of the Labelling GUI

To guarantee process stability in the labelling process, a matlab app is used. This application was designed specifically to ensure the manual labeling

process to be as precise as possible. A screenshot throughout the process can be seen in Figure 3. Therefore, a GUI, see in Figure 2, enables the data processor to mark the start and end position of the specific maneuver according to the timestamp of the video.The label for neutral is given to all data points that arise from normal riding influences without turns or jumps. To label the data, the matlab application $ValidationApp\_V1$ reads in the cut video file and plays it frame by frame. All the videos were recorded with a framerate of 30fps. In the end, the output of the labeling application consists of four labelled -csv files. Each of these contains one row of holding the time stamp, the X,Y,Z data of the according sensor and a column of the according label. The application further ensures that an event can only be started out of the neutral position. This means, when the button right turn is clicked, only the button to end the right turn can be clicked next.

Fig. 3. Screenshot of the Labelling application

For the usage of the labelled data set it is of high importance to ensure no mislabeled values. To minimize the risk of human error the labeled data set and especially the amount of turns and jumps is checked with a second view of the rides video.

C. Preprocessing for Neural Network

One essential step in working with neural networks is the preprocessing of the data. To do so, the data, as it contains more than one feature, has to be normalized. This helps the gradient descent converge more quickly. Further, it ensures that test values bigger than values from the training set are still in range of the network. And it helps the network to learn appropriate weights for the individual features, without giving more attention to features with a wider range [4].

To scale the data in the right range, the best approach would be to normalize the data in the range of the sensors capability. Due to the fact, that there is no information about the in mobile phones used sensors, the data has been checked for its range. The complete data set has been investigated regarding to the span of the values. This showed values between $\pm 80 m/s^2, \pm 80 rad/s, \pm 80 \mu T$. Therefore, the normalization is done in the following manner: 100 is added to all values to make them positive. Finally, to receive values in a positive range between 0 and 1, all values are divided by 200. This normalization step is carried out in a google colab script. Each scaling factor, in this case the scaling variables 100 and 200, can be adjusted separately for each input. In this case however, there was no need to do so. Figure 4 shows the signal of all sensors in a sequence of 18.000 data points. The last three graphs in this Figure show the binary onehotencoded label.

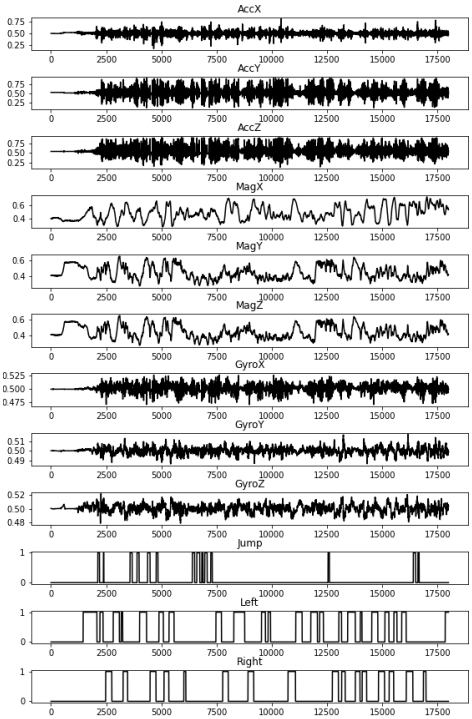


Fig. 4. Overview of all data

It is common practise for Neural Networks that the available dataset is split up. This is done due to the fact that the Neural Network has to be trained. And the trained Neural Network has to be evaluated. The Neural Networks uses a training set of data to organize its weights. When the weights are set, it checks these against the validation set. This is done over several loops to minimize the loss. The repetition of this training is called epoch. The training is finished when the loss does not decrease any more or an early stopping criteria is reached to prevent overfitting. When the Neural Network converges or stops, the trained network is checked for its performance with a previously not used data set, the so-called test set. In used ratio of training, to validation, to test set is 70%, 15% and 15% of the available labelled data. The data split is executed in a google colab script.

IV. Algorithm

With labelled data, a Neural Network can be trained to fulfill specific tasks. An algorithm solves this by learning from the data that is provided to it [5]. To classify turns, as well as jumps as the air time out of data, a neural network is designed. In the following, the structure and parameters of the network used in this paper, are described. Figure 5 explains the working principle of this algorithm. Using labelled data, a Neural Network is trained and able to predict on new sensor data. These predictions are evaluated with a tool called peakfinder.

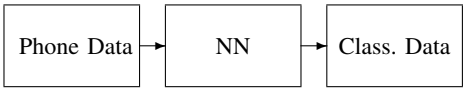


Fig. 5. Basic working principle of estimated system

Based on the structure and the use of CNNs and LSTM in one network, classifying human activity, in the work of Wang et. al. [**?**], a model out of convolutional and one LSTM layer is built. In consideration of the usage of acceleration data, this model is adapted to the case of this study. Additionally in the according thesis to this paper of the same author also the prediction of a CNN GRU model with the same parameters is evaluated.

The used model consists of CNN layers to extract features out of the sensor data, which are relevant for the specific maneuver. The LSTM layer is implemented to generate a memory over the whole data set to be able to generalize future data. The final, in colab implemented approach, can be seen in Figure 6.

```
Model: "Trailfeature_Detector"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
dense_6 (Dense)              (None, 100, 128)          1664

conv1d_2 (Conv1D)            (None, 69, 32)            131104

conv1d_3 (Conv1D)            (None, 62, 32)            8224

max_pooling1d_1 (MaxPooling  (None, 31, 32)            0
1D)

lstm_1 (LSTM)                (None, 32)                8320

dense_7 (Dense)              (None, 128)               4224

dropout_1 (Dropout)          (None, 128)               0

dense_8 (Dense)              (None, 64)                8256

dense_9 (Dense)              (None, 32)                2080

dense_10 (Dense)             (None, 8)                 264

dense_11 (Dense)             (None, 3)                 27

=================================================================
Total params: 164,163
Trainable params: 164,163
Non-trainable params: 0
```

Fig. 6. Model parameter

The implemented model has been optimized regarding its performance with the parameters of batch size and window length of the sliding window approach. In the end, a batch size of 1.375 and a window size of 100 or 200 is used. Further, the models performance to different configurations of sensor input was tested. The test's outcome is that the model works best when all sensors are used.

A. *Teaching Neural Network*

The previously described model has been trained with all sensors from the stem data. Stem data is the data that is recorded with the mobile phone mounted to the stem. Figure 7 shows that the model reached its maximum accuracy after 12 epochs. It has been stopped to train to prevent overfitting. It has to be stated that it more data would be needed to achieve more satisfying and precise results.

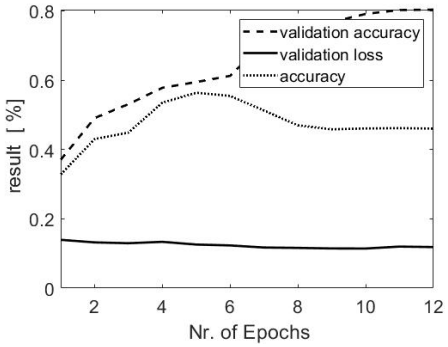


Fig. 7. Accuracy, Validation Accuracy and Validation loss over the epochs

The following confusion matrix, see Figure 8 shows a comprehensive version of the prediction of the model. As can be seen from this graphic, the model does not deliver satisfying values. Therefore, no useful prediction on air time, as well as the amount of turns can be evaluated.

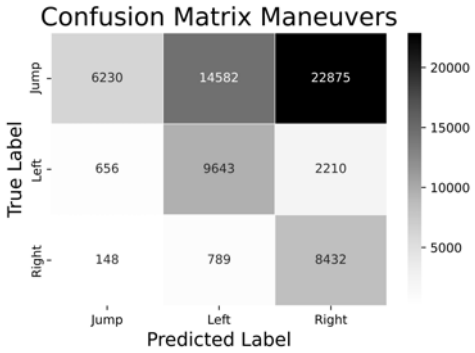


Fig. 8. Overview of all data

V. Results

This final chapter leads through all results of the author's work and provides an outlook into future possibilities, as well as thoughts and suggestions on areas to improve and develop further. Figure 9 shows

an overview of the prediction of the model on all data.

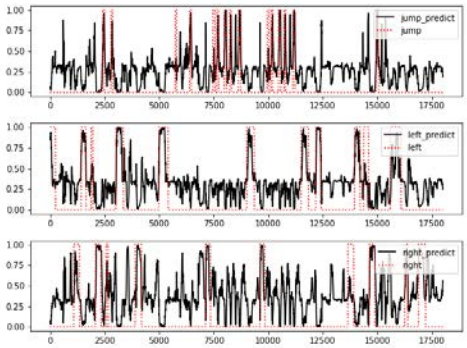


Fig. 9. Overview of prediction of all data

To see the performance of the model in each maneuver in a more detailed version, a zoomed in sequence is shown in the following three Figures 10, 11, 12.

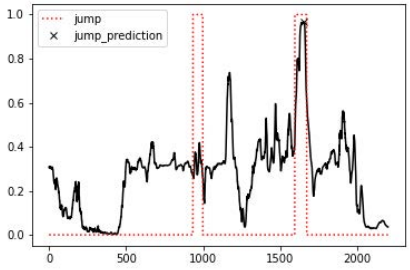


Fig. 10. Jump prediction

Figure 10 shows in black the prediction of the model in the sequence of the first 2.200 data points. The label is visible in red, this way the indication for the correct event is given. It is clearly visible, that there are jumps which are predicted very accurately by the model, which is the case for example on the peak seen on the right half of the figure. On the other hand, the jump on the left could not be identified with peakfinder due to its under average height of the other predictions of jumps. Additionally, in between these two peaks, the model suggests a jump where no jump is labelled. As seen from the graph, a clear determination of a jump and therefor, the estimation of the airtime is not possible. To determine the air time, the model's prediction has to delivery a prediction more likely to the right jump seen in Figure 10. Such results would make it possible to determine the edge peak points of the jump. Out of these, the air time would be calculated by the distance of the left and right data point.

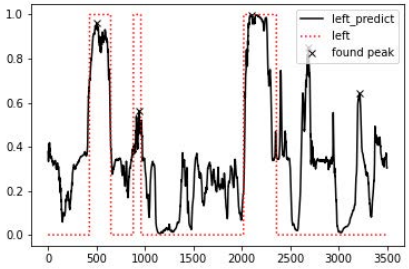


Fig. 11. Left turn prediction

To evaluate the quality of the prediction regarding left turns, the same procedure as for the jumps is applied. The parameters for the peakfinder have been determined out of Figure 11. It has to be stated, that the prediction regarding left turns is more accurate than the prediction on jumps. The prediction does not reach a satisfying level, but it does demonstrate a good indication of the network. Same as for the jumps, the prediction shows different peaks for different labels. There are peaks that fill up the label very well, but on the other hand, there are also peaks that do not point out as much as needed. With more data, the right peaks should develop a more precise characteristic and peaks outside the labelled zone should decrease. In case further training with

additional data would still generate predictions out of the labelled range, the labelled data would need to be rechecked for a missing label.

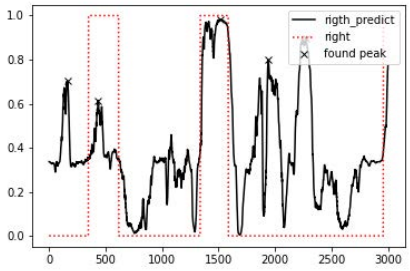


Fig. 12. Right turn prediction

To evaluate the prediction on the right turns, the same procedure as in the previous two sections is applied. Figure **??** shows the same characteristics as Figure 12. There are turns in the data set that are predicted quite accurately (turn around data point 1.500) and there are also turns without such good characteristics. Compare to turn around data point 500.

VI. Conclusion

Throughout the optimization process, the author recognized the tendency of the algorithm to optimize either on jumps or on turns. Concluding, to improve the algorithm's success, the whole algorithm should be split in two. One algorithm trained to detect jumps, and the other one trained upon detecting turns. Further, the algorithm trained on turns can potentially also be separated into right and left turns. Concluding from the discussion and the results, it can be said that a mobile phone is capable to record sensor data in the manner that different maneuvers can be classified. This thesis derives an algorithm, that shows the tendencies to do so. To be able of doing so in a manner that provides satisfying results, more data needs to be generated. Further, it is shown that generating and correct labelling of the data is absolutely essential for the quality of the algorithm and therefore a control loop needs to be implemented. At the same time, it again needs to be drawn into attention, that the data generation was and needs to be done in an as realistic setting as possible, in order to generate the most realistic data set. Whilst a simplified version of the ridden track could potentially increase the performance of the algorithm, it would at the same time not satisfy the aim of the development of a universal solution for mobile phones.

References

[1] "Mtbx," 04.05.2022. [Online]. Available: https://mtbx.bike//

[2] SAGE Journals, "Physical activity recognition utilizing the built-in kinematic sensors of a smartphone - yi he, ye li, 2013," 31.05.2022. [Online]. Available: https://journals.sagepub.com/doi/full/10.1155/2013/481580

[3] B. Mehlig, "Machine learning with neural networks," 2021. [Online]. Available: https://arxiv.org/pdf/1901.05639.pdf

[4] "Representation: Cleaning data — machine learning — google developers," 18.07.2022. [Online]. Available: https://developers.google.com/machine-learning/crash-course/representation/cleaning-data

[5] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016, http://www.deeplearningbook.org.

Maximilian Karl Freiheit is a passionate outdoor lover and wanted to connect his passion for technique with his passion for sport in this thesis.

Analyse eines Mikroschweißprozesses für MEMS-basierte Sensoren

Thomas Giuliani, Felix Gennrich (Betreuer) und Bernhard Hollaus (Gutachter)

***Kurzfassung*—Beschleunigungssensoren auf der Grundlage von mikro-elektromechanischen Systemen (MEMS) sind in der Unterhaltungsindustrie sowie der Automobilindustrie weit verbreitet. Um die Empfindlichkeit solcher MEMS pro Fläche zu erhöhen und damit die Baugröße zu verringern, wird derzeit ein Verfahren zur nachträglichen Spaltreduktion entwickelt. Dieser Prozess umfasst das Schweißen von zwei Siliziumverbindungen über Aluminiumleiterbahnen. Die aktuelle Prozessführung führt dazu, dass Material von den Aluminiumleiterbahnen in die Schweißregionen fließt und die Silizium-Verbindung im Schweißpunkt kontaminiert, was die mechanische Stabilität drastisch reduziert. Im Rahmen dieser Arbeit soll der Schweißprozess im Detail analysiert und ein vereinfachtes Modell zur Beschreibung des Mikroschweißprozesses aufgestellt werden. Es gilt Verbesserungsvorschläge für den Schweißprozess sowie die Schweißgeometrie auszuarbeiten und die Grundlage für weitere simulative Untersuchungen zu schaffen.**

Durch die Messung der U-I-Kennlinie wird das Widerstandsverhalten der Schweißstrecke vor dem Verschweißen bestimmt. In einer Literaturrecherche konzentriert sich der Autor auf die thermischen bzw. elektrischen Eigenschaften von Silizium und Aluminium. Im weiteren Verlauf wird die betrachtete Schweißstrecke diskretisiert und ein verhältnismäßig einfaches Modell in Simulink entwickelt. Die Validierung des vorgeschlagenen Modells umfasst den Widerstandsverlauf während des Schweißprozesses. Zur Bewertung der erzielten Schweißergebnisse werden Rasterelektronenmikroskopie (REM) und energiedispersive Röntgenspektroskopie (EDX) eingesetzt.

Während des Schweißpulses stimmen die Simulationsergebnisse gut mit den Messergebnissen überein, wobei die Dynamik der Messung etwas höher ist. Messungen mit REM und EDX bestätigen gute Schweißergebnisse mit dem optimierten Schweißprozess, obwohl ein Aluminiumfluss in die Schweißregionen nicht vollständig verhindert werden konnte. Darüber hinaus werden auf Grundlage der gewonnenen Ergebnisse einige Verbesserungsvorschläge aufgegriffen.

***Schlagwörter*—Mikroschweißen, Silizium, Modellierung, MEMS, Beschleunigungssensor**

T. Giuliani studiert am Studiengang Mechatronik, MCI, Innsbruck, Österreich, e-mail: thomas.giuliani@gmx.net.

F. Gennrich arbeitet bei MED-EL, Innsbruck, Österreich.

B. Hollaus arbeitet am Studiengang Medizintechnik, MCI, Innsbruck Österreich.

Manuskript eingereicht am 13. Mai 2022; revidiert am 30. Mai 2022.

SYMBOLVERZEICHNIS

A	Querschnittsfläche	$\mu\mathrm{m}^2$
b	Breite	μm
C_p	Wärmekapazität	$\mathrm{J\,mol^{-1}\,K^{-1}}$
d	Schichtdicke	nm
E_m	Schmelzenergie	$\mathrm{kJ\,mol^{-1}}$
λ	Thermische Leitfähigkeit	$\mathrm{W\,m^{-1}\,K^{-1}}$
l	Länge	μm
p_MEMS	MEMS Innendruck	Pa
σ_Al	Spezifischer Al-Widerstand	$\mathrm{S\,m^{-1}}$
σ_Si	Spezifischer Si-Widerstand	$\mathrm{S\,m^{-1}}$
T	Temperatur	K
T_amb	Umgebungstemperatur	K
$T_\mathrm{m\_Al}$	Al-Schmelztemperatur	K
$T_\mathrm{m\_Si}$	Si-Schmelztemperatur	K
U	Elektrische Spannung	V

ABKÜRZUNGSVERZEICHNIS

Al	Aluminium
EDX	Energiedispersive Röntgenspektroskopie
FEM	Finite-Elemente-Methode
MEMS	Mikro-elektromechanisches System
MSP	Mikroschweißprozess
REM	Rasterelektronenmikroskopie
Si	Silizium
SP	Schweißpunkt

I. Einleitung

Die Entwicklung des ersten mikroelektromechanischen Systems (MEMS) geht auf das Jahr 1967 zurück, wobei der erste ausgereifte Beschleunigungssensor 1979 entwickelt und getestet wurde [1], [2]. Heutzutage sind MEMS-Sensoren aufgrund ihrer hohen Sensitivität bei geringer Größe und Gewicht fester Bestandteil unseres täglichen Lebens und finden beispielsweise in der Automobilindustrie sowie Unterhaltungsindustrie (in z.B. Smartphones) Anwendung [3], [4], [5].

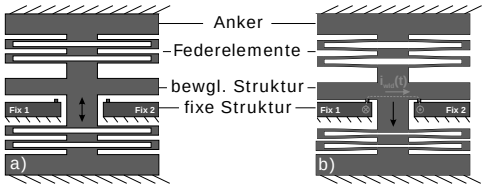


Abbildung 1. Der MSP schematisch dargestellt [6]. a) Initiale Position. b) Während dem Schweißprozess.

Ein Forschungsbereich für die Weiterentwicklung von MEMS betrifft das Erhöhen der Sensitivtät pro Fläche. Da das übliche Verfahren zur MEMS Fertigung eine Limitierung hinsichtlich des Aspektverhältnisses aufweist, soll ein nachträglicher Prozessschritt eingeführt werden, der eine Erhöhung der Sensitivität verspricht. Mittels Feldspannung wird der Spaltabstand reduziert und per Mikroschweißprozess (MSP) fixiert. Damit ist es möglich, die Grundkapazität des MEMS Differentialkondensators zu erhöhen und folglich auch dessen Sensitivität [6], [7], [8]. Eine schematische Darstellung dieses Vorgangs ist in Abbildung 1 ersichtlich. Der Innendruck der betrachteten MEMS liegt bei $p_{\text{MEMS}} = 5\,\text{mBar}$.

Da für einen industriell tauglichen Prozess, im Gegensatz zu den Forschungsarbeiten [6], [7], [8], das Einbringen des Schweißstroms nicht direkt an den Schweißpunkten (SP) erfolgen kann, ist die Verwendung von Zuleitungen erforderlich. Hierfür kommen oberflächlich aufgebrachte Al-Leiterbahnen zum Einsatz. Dies birgt die Gefahr, dass es unter Anwendung des MSP, wie er in [6] beschrieben wird, zum Aufschmelzen der Al-Leiterbahnen kommt. Im Extremfall kann es, wie in Abbildung 2 ersichtlich, durch den MSP zur Ausbildung einer Al-Brücke kommen, die allerdings nur eine begrenzte mechanische Festigkeit verspricht.

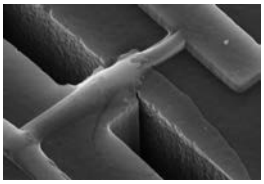


Abbildung 2. REM-Aufnahmen eines Mikroschweißkontaktes mit ausgebildeter Al-Brücke. Die hellgraue Struktur zeigt die Al-Leiterbahnen. Die dunklere Struktur stellt das Grundmaterial, den Si-Einkristall, dar

Im Rahmen dieser Arbeit soll der MSP im Detail analysiert und simulativ untersucht werden. Hierzu wird das verwendete Material (Al, Si) über einen Temperaturbereich von $300\,\text{K}$ bis über den Schmelzpunkt hinaus modelliert und ein vereinfachtes Modell in der Entwicklungsumgebung *Simulink* erstellt.

II. Systemanalyse & Modellbildung

Im Folgenden wird die Vorgehensweise bei der Erstellung des Simulink-Modells betreffend die Schweißstrecke, die Materialeigenschaften und das Simulationsmodell beschrieben.

A. Die Schweißstrecke

Das Layout der Schweißstrecke der untersuchten MEMS wird in Abbildung 3 gezeigt. Infolge der eingebrachten elektrostatischen Spannung lenkt die bewegliche Struktur in Richtung der fixen Struktur aus. Somit wird an den beiden SP ein elektrischer Kontakt hergestellt.

B. Materialeigenschaften

Da im Bereich des Schmelzpunktes von Al und Si sprunghafte Änderungen im Materialverhalten auftreten, steigt der numerische Aufwand für die Berechnung des transienten Modells so stark an, dass in

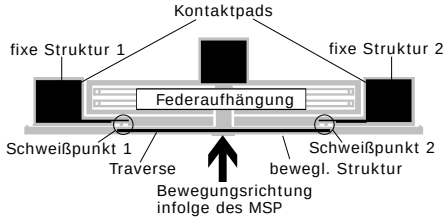


Abbildung 3. Layout der Schweißstrecke. Die hellgraue Struktur besteht aus Si-Einkristall. Die schwarze Struktur zeigt die Leiterbahnen auf Al-Basis. Die Kontaktpads dienen zum Einbringen des Schweißstroms. Die bewegliche Struktur ist über eine Feder aufgehängt.

diesem Bereich das Modell nicht gelöst werden kann. Daher wird das Materialverhalten in diesem Bereich, wo es notwendig ist, linearisiert. Der Verlauf des spezifischen elektrischen Leitwerts von Si mit einer Linearisierung von $\pm 70\,\mathrm{K}$ um den Schmelzpunkt $T_{\mathrm{m\_Si}} = 1687\,\mathrm{K}$ [9] findet sich in Abbildung 4.

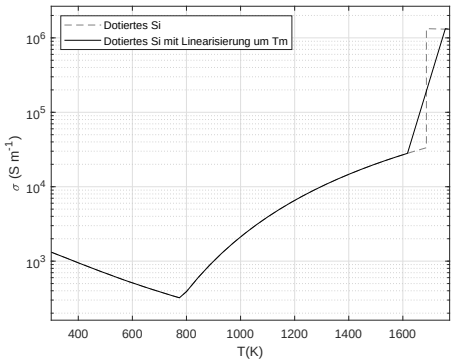


Abbildung 4. Elektrischer Leitwert von Si von 300 K bis 1800 K basierend auf den Daten von [10], [11], [12]. Der intrinsische Bereich von Si wird bei etwa 800 K erreicht.

In Tabelle I finden sich die Werte des konduktiven Leitwert von Si linearisiert mit $\pm 30\,\mathrm{K}$ um den Schmelzpunkt. Nach [13], [14] ist der radiative Anteil bis hin zum Schmelzpunkt vernachlässigbar, da dieser maximal etwa 2 % des Gesamtanteils der thermischen Leitfähigkeit ausmacht. Der konvektive Term wird aufgrund des sehr geringen Innendrucks vernachlässigt.

Tabelle I
KONDUKTIVER THERMISCHER LEITWERT VON SI [9].

T (K)	λ (W m^{-1} K^{-1})	T	λ
293	139,4	1173	24,8
373	105,8	1273	23,7
473	82,4	1373	22,3
573	65,6	1473	20,1
673	50,3	1573	20,4
773	42,4	1657	19,2
873	36,7	1717	57,0
973	30,9	1724	57,4
1073	27,4		

Weiters wird die Schmelzenthalpie von Si nach [15] mit $E_{\mathrm{m}} = 50{,}66\,\mathrm{kJ\,mol^{-1}}$ angegeben und über die Wärmekapazität im Temperaturbereich von $1687\,\mathrm{K} \pm 70\,\mathrm{K}$ modelliert. Nach [16] kann die Wärmekapazität mit

$$C_{\mathrm{p}} = 23{,}698 + 3{,}305 \cdot 10^{-3}\,T - 4{,}354 \cdot 10^{5}\,T^{-2} \quad (1)$$

definiert werden.

Die elektrische Leitfähigkeit bzw. die Wärmekapazität von Al finden sich linearisiert mit $\pm\,70\,\mathrm{K}$ um die Schmelztemperatur $T_{\mathrm{m\_Al}} = 933\,\mathrm{K}$ in Tabelle II und III [17]. Die thermische Leitfähigkeit wird nach [18] über das Wiedemann-Franz Gesetz

$$\lambda = \sigma_{\mathrm{Al}}\,L\,T \quad (2)$$

von der elektrischen Leitfähigkeit abgeleitet, wobei $L \approx 2{,}445\,10^{-8}\,\mathrm{V^2\,K^{-2}}$ die Lorenz-Zahl definiert.

Die Grenzschicht im SP wird durch einen geschalteten elektrischen Widerstand modelliert und so gewählt, dass er über die Schweißstrecke bei einer umgesetzten Leistung von 300 mW bei $7\,\mathrm{k\Omega}$ (vgl. Abschnitt IV-A) liegt. Der Widerstand wird beim Erreichen von $T = T_{\mathrm{m\_Si}}$ kurzgeschlossen und bleibt folglich in diesem Zustand. Nach dem Erreichen der Schmelztemperatur werden die fixe Struktur und die bewegliche Struktur im Modell thermisch verbunden.

Tabelle II
SPEZIFISCHER ELEKTRISCHER LEITWERT VON AL [18].

T (K)	σ (S m^{-1})	T	σ
298	36,9·10^6	773	11,6·10^6
384	30,5·10^6	823	10,8·10^6
373	27,5·10^6	863	9,0·10^6
473	20,3·10^6	1003	4,1·10^6
573	16,2·10^6	1023	3,9·10^6
673	13,7·10^6		

Tabelle III
WÄRMEKAPAZITÄT VON AL [19].

T (K)	C_p (J mol^{-1} K^{-1})	T	C_p
298	5,8	700	7,0
300	5,8	800	7,4
400	6,2	863	7,7
500	6,5	1003	7,6
600	6,7	1700	7,6

Die Übergangsschicht zwischen Al und Si wird als sehr dünn eingeschätzt, da in den elektrischen Messungen kein ausgeprägter p-n-Übergang nachgewiesen werden konnte. Der eindimensionale thermische Widerstand, wie er bei den betrachteten Strukturen beim Übergang von Al zu Si angenommen werden kann, wird in [20] mit

$$R_\mathrm{L} = \frac{d}{\lambda\, A} \tag{3}$$

definiert. d entspricht der Schichtdicke, A der Fläche und λ der Wärmeleitfähigkeit. Da die Schichtdicke als sehr gering und die Fläche im Vergleich dazu als relativ groß angenommen werden kann, wird der resultierende thermische Widerstand klein und daher vernachlässigt.

C. Energiequelle

Nach [21] gibt es für einen Schweißprozess stets eine obere und eine untere Grenze der Leistungsdichte. Daher liegt der Versuch nahe, die eingebrachte Schweißleistung zu kontrollieren. Folglich wird zum Schweißen eine Leistungsquelle verwendet, die durch das gezielte Schalten von Vorwiderständen und der Spannungsversorgung generiert wird.

D. Diskretisierung

Die Diskretisierung der Geometrie erfolgt im Bereich um die SP in x- und y-Richtung mit relativ hoher Auflösung, da dieser Bereich im Detail untersucht wird. Für die Schweißstrecke kann eine Symmetrieebene zwischen den fixen Strukturen und der beweglichen Struktur entlang des in Abbildung 3 eingezeichneten Pfeils gefunden werden. Daher wird nur ein Teil der Schweißstrecke modelliert und somit der numerische Aufwand reduziert. Die Zuleitungsstrecken werden mit vergleichsweise geringer Auflösung modelliert. Das gesamte Modell ist in Abbildung 5 dargestellt.

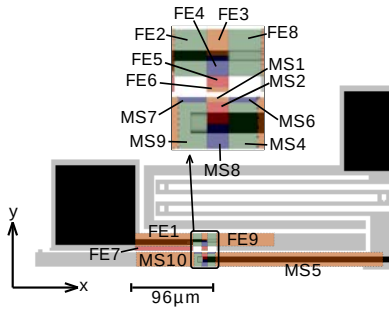


Abbildung 5. Diskretisierung der Schweißgeometrie. Die grauen Strukturen bestehen aus Si-Einkristall. Die schwarzen Strukturen bilden die Al-Leiterbahnen.

In z-Richtung erfolgt die Modellierung in vier Schichten, die aus praktischen Überlegungen heraus wie folgt gewählt wurden:

1) Die Al-Leiterbahnen weisen eine Dicke von etwa $d_\mathrm{Al} = 1{,}5\,\mu$m auf. Die Leiterbahnen werden thermisch sowie elektrisch modelliert.
2) Die erste Schicht im Si wird mit einer Dicke von $d_\mathrm{Si_1} = 3\,\mu$m definiert. Sie wird thermisch und elektrisch modelliert.
3) Die zweite Schicht mit einer definierten Dicke von $d_\mathrm{Si_2} = 17\,\mu$m wird rein thermisch betrachtet.
4) Die dritte Schicht mit einer definierten Dicke von $d_\mathrm{Si_3} = 55\,\mu$m an den beweglichen

Strukturen bzw. $d_{\text{Si\_3}} = 65\,\mu\text{m}$ an den fixen Strukturen wird rein thermisch betrachtet (siehe Abbildung 3).

E. Gesamtmodell

Im Gesamtmodell wird, wie bereits in Abschnitt II-D erläutert, nur die halbe Schweißstrecke modelliert. Das elektrische Modell ist schematisch in Abbildung 6 ersichtlich. Das thermische Grundmodell wird in Abbildung 7 gezeigt. Jedes modellierte Element besitzt eine Wärmekapazität sowie n Eingänge und m Ausgänge, die es mit den jeweiligen benachbarten Elementen verbindet. Die elektrischen Widerstände sowie die thermischen Eigenschaften der modellierten Elemente sind temperaturabhängig und dadurch mit der Temperatur gekoppelt, welche in der Wärmekapazität des jeweiligen Elements gespeichert ist.

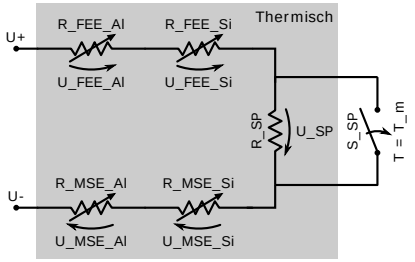


Abbildung 6. Elektrisches Schweißmodell im Überblick. Der SP wird elektrisch durch einen Schalter modelliert. Die thermische Modellierung erfolgt für alle Elemente im betrachteten Modell. Elektrisch werden nur die Al-Leiterbahnen und die erste Si-Schicht, die zum elektrischen Stromfluss beiträgt, modelliert.

Das Gesamtmodell besteht aus der modellierten Schweißstrecke und einer Leistungsquelle, die einen Leistungsverlauf nach Abbildung 8 generiert.

Die Randbedingungen für das beschriebene Modell werden wie folgt definiert:

- Die Modellränder, welche an die umliegende Niederdruck-Atmosphäre grenzen, werden als adiabatische Wände definiert.

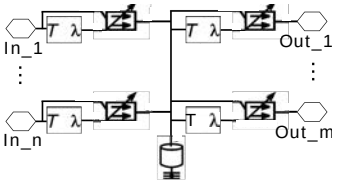


Abbildung 7. Struktur des thermischen Modells für ein einzelnes diskretes Element.

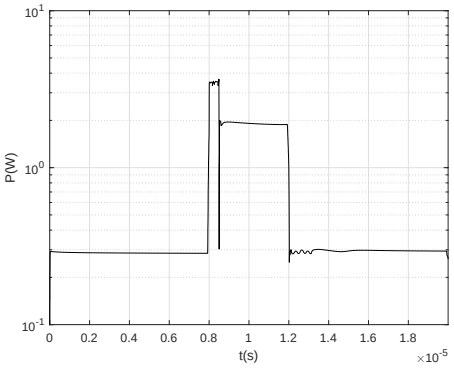


Abbildung 8. Leistungsverlauf über die gesamte Schweißstrecke in der Simulation. Die Daten werden mittels gleitendem Mittelwert über 35 Datenwerte gemittelt. Somit können Flanken aufgrund der Schaltvorgänge reduziert werden.

- Die Modellränder, welche an das umliegende Si angrenzen, werden mit $T_{\text{amb}} = 300\,\text{K}$ angenommen.

III. Messaufbau & Schweisselektronik

Zur Validierung des Simulationsmodells werden Messungen an realen Komponenten vorgenommen, welche durch den optimierten MSP geschweißt werden. Der Aufbau zu den praktischen Untersuchungen findet sich in Abbildung 9. Hierbei wird die Schweißelektronik mit einer Leistungsquelle von $+V_{\text{cc}}$ nach Masse versorgt. Über einen Shuntwiderstand wird der zugeführte Strom bestimmt. Dafür kommt für den statischen Fall ein Multimeter zum Einsatz. Im

dynamischen Fall werden die Messergebnisse über ein Oszilloskop mit Differentialtastkopf generiert. Der Spannungsabfall über die gesamte Schweißstrecke wird ebenfalls dynamisch über das Oszilloskop gemessen und protokolliert. Für die Schweißung und die Messung wurde das gelistete Equipment aus Tabelle IV verwendet.

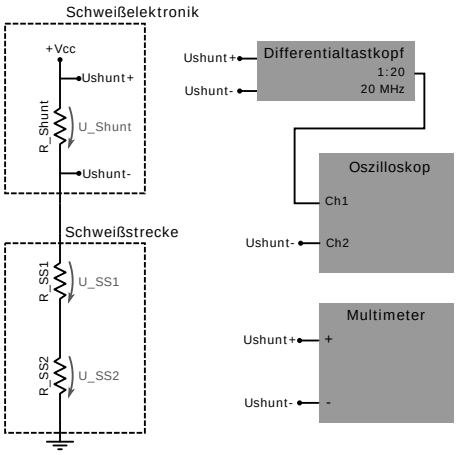


Abbildung 9. Schweißelektronik und Messaufbau.

Tabelle IV
EQUIPMENT FÜR DIE PRAKTISCHEN UNTERSUCHUNGEN

Gerät	Hersteller	Bezeichnung
Oszilloskop	RIGOL	DS1054Z
Tastkopf	TESTEC	TT-SI 9002
Multimeter	KEITHLEY	2000
Prober	Cascade	PA200 "BlueRay
LCR-Meter	Keysight	E4980AL

IV. ERGEBNISSE

A. *Experimentelle Schweißergebnisse*

Um das Verhalten der Schweißstrecke an mehreren Proben zu untersuchen, wurde zunächst die in Abbildung 10 dargestellte U-I-Kennlinie aufgenommen. Es zeigt sich ein sehr reproduzierbares Verhalten bei einer zugeführten Leistung von 300 mW. Der Widerstand liegt bei etwa 7 kΩ vor dem Verschweißen.

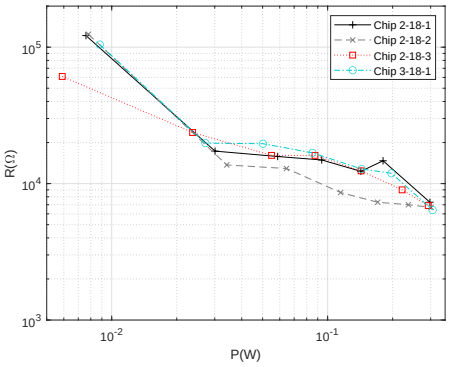


Abbildung 10. Strom-/Spannungsmessung zur Voruntersuchung an vier zufälligen Proben an einem Wafer. Über die gesamte Untersuchung wird die bewegliche Struktur mittels elektrostatischer Kraft an die fixe Struktur gepresst.

Während der Schweißung zeigt sich für die Proben in Abbildung 11 wiederum ein sehr reproduzierbares elektrisches Verhalten. Die Oszilloskop-bedingte Messung am Shuntwiderstand ist bei hohen Widerständen im kΩ-Bereich deutlich verrauscht, weswegen die initialen Widerstandswerte bei dieser Messung deutlich von den zuvor bestimmten 7 kΩ abweichen. Infolge des Schweißimpulses sinkt der Gesamtwiderstand innerhalb von 0,12 μs um über zwei Größenordnungen. Nach dem Impuls steigt der Widerstand aufgrund der Abkühlung im Si stark an und liegt für die untersuchten MEMS am Ende bei etwa 200 Ω.

Abbildung 12 zeigt die REM-Aufnahme einer Schweißung, die im Zuge der praktischen Untersuchungen durchgeführt wurde. Zu sehen ist, dass es an der Verjüngung der Al-Leiterbahn zu augenscheinlichem Al-Fluss gekommen ist. Daher ist auch der niedere Widerstand nach der Schweißung erklärbar. Nach analytischen Berechnungen sollte der Widerstand über die Schweißstrecke ohne Al-Beteiligung

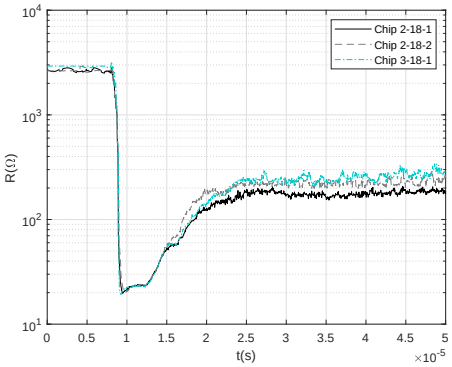


Abbildung 11. Widerstandsverlauf in den praktischen Untersuchungen. Die Daten werden mittels gleitendem Mittelwert über 10 Datenwerte geglättet.

bei etwa $1\,\mathrm{k\Omega} - 2\,\mathrm{k\Omega}$ liegen. Mittels EDX-Analyse konnte nachgewiesen werden, dass der Al-Gehalt in der Nähe des Schweißpunktes bei etwa 6 % liegt und somit deutlich geringer ist als bei der ursprünglichen Prozessführung.

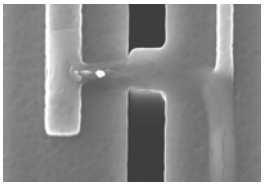


Abbildung 12. REM-Aufnahmen eines erfolgreich geschweißten MEMS. Die hellgrauen Strukturen stellen die Al-Leiterbahnen dar. Die dunkelgrauen Strukturen bestehen aus Si-Einkristall.

B. Simulationsergebnisse

In Abbildung 13 werden die simulativ bestimmten Temperaturverläufe in ausgewählten Si Elementen dargestellt. Im Verlauf zeigt sich, dass am Anfang des Impulses besonders die beiden Elemente FE6_1 und MS1_1 aufgeheizt werden. Dies ist insofern nachvollziehbar, da der Schweißpunktwiderstand maßgeblich ist. Nach dem Erreichen der Temperaturspitze wird der Schweißpunktwiderstand kurzgeschlossen. Dadurch sinkt die in diesem Bereich abfallende Spannung und damit auch die generierte Leistung deutlich ab. Mit fortschreitender Impulsdauer gleichen sich die benachbarten Elemente FE6_1 und FE5_1 bzw. MS1_1 und MS2_1 immer näher an.

Wenn man nun jedoch den Fokus auf die beiden Elemente FE6_2 und MS1_2 legt, so zeigt sich ein gravierender Temperatursprung zwischen der ersten und zweiten Schicht. Dies deutet daraufhin, dass das thermische Verhalten nur unzureichend abgebildet werden kann.

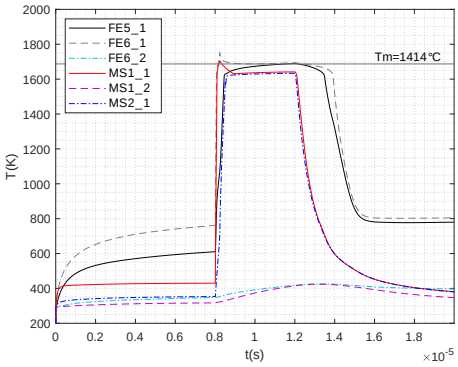


Abbildung 13. Temperaturverläufe im Si simulativ bestimmt. Die genaue Beschreibung zu den simulierten Elementen findet sich in Abschnitt II-D.

In Abbildung 14 wird der Widerstandsverlauf der simulierten Schweißstrecke dargestellt. Der Widerstand nimmt infolge des Schweißimpulses um über zwei Größenordnungen ab. Der Widerstand im Si bleibt jedoch über den gesamten Schweißprozess deutlich größer als jener im Al und ist daher maßgeblich für den Gesamtwiderstand.

V. Diskussion

Um der sehr hohen Änderungsrate des Si-Widerstandes zu begegnen, ist die Ansteuerung mit-

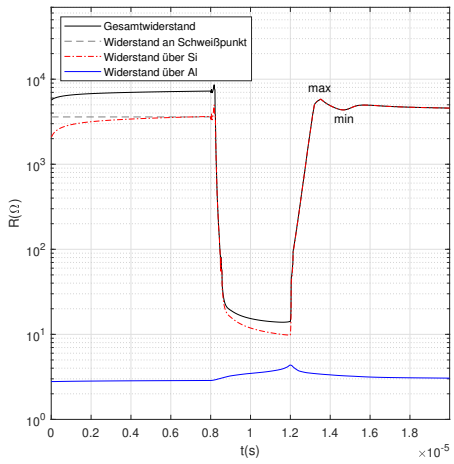


Abbildung 14. Widerstandsverlauf simulativ bestimmt. *Min* und *Max* beschreiben hierbei das lokale Minimum und Maximum des Widerstandes nach dem Abkühlen des Si.

tels Leistungsquelle die beste Wahl. Damit ist es möglich, einen entsprechenden Parametersatz zu definieren, mit dem ein reproduzierbarer Prozess gefahren werden kann. Im MSP selbst kann beispielsweise der Widerstandsverlauf als Indikator zum Einsatz kommen, ob eine Schweißung erfolgreich ist. In der Modellvalidierung konnte gezeigt werden, dass mit dem relativ einfachen vorgeschlagenen Modell der MSP in seiner Grundcharakteristik abgebildet werden kann. Allerdings wurde auch deutlich, dass es für eine genaue Systembeschreibung einer Modellierung auf z.B. FEM Basis bedarf. Besonders das thermische Modell weist hinsichtlich der Temperaturverläufe entsprechende Schwachstellen auf - die Dynamik wird unterschätzt.

Um feststellen zu können, ob es während dem MSP zum Al-Fluss in die Schweißstellen gekommen ist, erweist sich die Impedanzmessung als Mittel der Wahl. Besonders niedere Realteile mit etwa $300\,\Omega$ und weniger weisen auf einen solchen Fluss hin. Prinzipiell ist es wünschenswert, dass die SP einen möglichst hohen Anteil an Si aufweisen, da hier von einer sehr steifen Verbindung ausgegangen werden kann, die nur sehr wenig Bewegung infolge von Beschleunigungen entwickelt.

Als alternative Möglichkeit zum Widerstandsschweißprozess kann beispielsweise das Laserschweißen gesehen werden. Hierbei wird von [22] berichtet, dass die Laserschweißmethode die günstigste Möglichkeit sei, Schweißungen im kleinen Maßstab durchzuführen. Der große Vorteil des vorgestellten Schweißverfahrens besteht jedoch darin, dass der MSP als eigenständiger Prozess nach vollständiger Waferfertigung angewendet werden kann.

VI. Schlussfolgerung

Aus den gewonnenen Erkenntnissen kann einerseits die Notwendigkeit einer Verbesserung der Leiterbahgeometrie zur Vermeidung von Kontamination durch Al im SP abgeleitet werden. Hierfür sollte das Layout so angepasst werden, dass die Al-Leiterbahnen nicht zu nahe am SP liegen. Andererseits wurde deutlich, dass breitere Leiterbahnen aufgrund der höheren Wärmekapazität und besseren thermischen Leitung weniger aufschmelzen als schmale Leiterbahnen.

Um den MSP simulativ besser beschreiben zu können, wird eine Simulation auf z.B. FEM Basis nötig sein. Insbesondere gilt es zu kontrollieren, dass Temperatursprünge zwischen einzelnen Elementen vermieden werden. Hierfür wird besonders im Schweißbereich eine sehr feine Granularität des Gitters nötig sein.

Im Rahmen der Forschungsarbeit konnte ein tiefes Verständnis über den MSP erlangt werden. Es konnte gezeigt werden, dass eine Leistungsquelle eine gute Ansteuerungsmöglichkeit für einen MSP im Si bildet. Für nachfolgende FEM-Untersuchungen ist beispielsweise die Bestimmung des Strompfades interessant und kann somit zur Optimierung der Schweißgeometrie herangezogen werden.

LITERATUR

[1] H. Muro, "History and recent progress of MEMS physical sensors," *Advances in science and technology*, vol. 81, pp. 1–8, 2013.

[2] L. M. Roylance and J. B. Angell, "A batch-fabricated silicon accelerometer," *IEEE Transactions on Electron Devices*, vol. 26, no. 12, pp. 1911–1917, 1979.

[3] S. Finkbeiner, "Mems for automotive and consumer electronics," in *2013 Proceedings of the ESSCIRC (ESSCIRC)*. IEEE, 2013, pp. 9–14.

[4] A. Sethuramalingam, TK und Vimalajuliet, "Design of MEMS based capacitive accelerometer," in *2010 International Conference on Mechanical and Electrical Technology*. IEEE, 2010, pp. 565–568.

[5] R. Bogue, "Recent developments in MEMS sensors: A review of applications, markets and technologies," *Sensor Review*, 2013.

[6] M. Haubold, "Erarbeitung einer Fertigungstechnologie und Charakterisierungsmethodik für die Herstellung hochsensitiver Vibrationssensoren unter Nutzung des Mikroschweißprozesses," Ph.D. dissertation, Technische Universität Chemnitz, 2016.

[7] M. Nowack, S. Leidich, D. Reuter, S. Kurth, M. Küchler, A. Bertz, and T. Gessner, "Novel post-process gap reduction technology of high aspect ratio microstructures ultilizing micro welding," in *2011 16th International Solid-State Sensors, Actuators and Microsystems Conference*. IEEE, 2011, pp. 1352–1355.

[8] M. Nowack, S. Leidich, D. Reuter, S. Kurth, M. Kuechler, A. Bertz, and T. Gessner, "Micro arc welding for electrode gap reduction of high aspect ratio microstructures," *Sensors and Actuators A: Physical*, vol. 188, pp. 495–502, 2012.

[9] E. Yamasue, M. Susa, H. Fukuyama, and K. Nagata, "Thermal conductivities of silicon and germanium in solid and liquid states measured by non-stationary hot wire method with silica coated probe," *Journal of crystal growth*, vol. 234, no. 1, pp. 121–131, 2002.

[10] S. Reggiani, M. Valdinoci, L. Colalongo, M. Rudan, G. Baccarani, A. D. Stricker, F. Illien, N. Felber, W. Fichtner, and L. Zullino, "Electron and hole mobility in silicon at large operating temperatures." *IEEE Transactions on Electron devices*, vol. 49, no. 3, pp. 490–499, 2002.

[11] S. Fan, G. Plascencia, and T. Utigard, "High temperature electric conductivity of pure silicon," *Canadian Metallurgical Quarterly*, vol. 47, no. 4, pp. 509–512, 2008.

[12] H. Schnyders and J. Van Zytveld, "Electrical resistivity and thermopower of liquid Ge and Si," *Journal of Physics: Condensed Matter*, vol. 8, no. 50, p. 10875, 1996.

[13] C. J. Glassbrenner and G. A. Slack, "Thermal conductivity of silicon and germanium from 3 K to the melting point," *Physical Review*, vol. 134, no. 4A, p. A1058, 1964.

[14] H. Shanks, P. Maycock, P. Sidles, and G. Danielson, "Thermal conductivity of silicon from 300 to 1400 k," *Physical Review*, vol. 130, no. 5, p. 1743, 1963.

[15] H. Sicius, "Kohlenstoffgruppe: Elemente der vierten Hauptgruppe," in *Handbuch der chemischen Elemente*. Springer, 2021, pp. 211–276.

[16] V. Glazov and A. S. Pashinkin, "The thermophysical properties (heat capacity and thermal expansion) of single-crystal silicon," *High Temperature*, vol. 39, no. 3, pp. 413–419, 2001.

[17] S. I. Bakhtiyarov, R. A. Overfelt, and S. G. Teodorescu, "Electrical and thermal conductivity of a319 and a356 aluminum alloys," *Journal of Materials Science*, vol. 36, no. 19, pp. 4643–4648, 2001.

[18] R. Brandt and G. Neuer, "Electrical resistivity and thermal conductivity of pure aluminum and aluminum alloys up to and above the melting temperature," *International Journal of Thermophysics*, vol. 28, no. 5, pp. 1429–1446, 2007.

[19] R. A. McDonald, "Enthalpy, heat capacity, and heat of fusion of aluminum from 366. degree. to 1647. degree. k," *Journal of Chemical and Engineering Data*, vol. 12, no. 1, pp. 115–118, 1967.

[20] E. Macherauch and H.-W. Zoch, *Praktikum in Werkstoffkunde: 100 ausführliche Versuche aus wichtigen Gebieten der Werkstofftechnik*. Springer-Verlag, 2019.

[21] J. F. Lancaster, "The physics of welding," *Physics in technology*, vol. 15, no. 2, p. 73, 1984.

[22] S. S. Gajapathi, "Heat transfer analysis of microwelding using tuned electron beam," Master's thesis, University of Alberta, 2011.

Thomas Giuliani ist Student am Masterstudiengang für Mechatronik am MCI Innsbruck/Österreich. Zudem ist er bei MED-EL als Entwickler im Fachbereich Mikrosystemtechnik tätig; hier insbesondere in der Entwicklung und Optimierung von MEMS.

Evaluation of Electrical Influences on Li-ion Batteries with Alternative Cell Arrangement in Parallel Connection

Markus Hausladen, Christoph Reiter (supervisor), and Bernhard Hollaus (supervisor)

***Abstract*—Developing durable and high-performance Li-ion battery systems is a challenge. However, the requirements are similar for many applications: Maximum performance with minimum package volume. This can be achieved by an unsymmetrical arrangement of cells. At high performance and poor design, there is a risk of increased cell aging.**

Therefore, a literature review on the effects of battery aging and interconnection concepts is done. Furthermore, a simulation model is used to ensure the battery performance. Finally, the battery simulation model is also validated by test results. The test setup is based on a minimal example of a battery for an electric motorcycle. In addition, the test results are used to optimize the development process for cells and batteries. A load cycle corresponding to that of an electric motorcycle is selected for the battery. This test concept ensures that the cells in the battery pack are loaded similarly to those in a real motorcycle battery.

In addition, the simulation model is validated using two different methods (electrical impedance spectroscopy and the pulse method). The simulation results of both measurement methods are compared with each other. Thus, it is ensured to use the correct method for the simulation of a load cycle in the future. In this way, the methods developed (simulation and test by minimum test setup) ensure at an early stage that unsymmetrically built battery systems achieve the required service life. Furthermore, no expensive battery prototype is necessary to draw conclusions about the performance of the battery.

***Index Terms*—Li-ion battery, Cell arrangement, Battery simulation, Electromobility, Parallel connection, Internal resistance, EIS measurement;**

M. Hausladen study at MCI,
e-mail: hausladen.markus@web.de

I. Introduction

AN unsymmetrically connected battery offers a number of advantages, especially since the available installation space for batteries in motorcycles can be optimally utilized. In this work, an unsymmetrical motorcycle battery concept is considered that allows a 14s10p interconnection, while a conventional design with symmetrical cell arrangement would only allow an interconnection of 14s8p. This corresponds to a higher capacity of 140 Ah in the same installation space with a 5 Ah cylindrical cell. The disadvantage of the unsymmetrical battery concept is that free-space cell connectors cause an unsymmetrical load on the cells in the parallel connection. Whether this unsymmetrical cell arrangement has major disadvantages in terms of unsymmetrical cell aging and battery performance remains to be investigated. [1], [2]

For this purpose, a possible test setup and measurement methods for the cell behavior will be examined in more detail. Furthermore this test results can be also used in simulation models. Thermal influences on the test setup can also be optimized and analyzed during test operation by temperature sensors. However, the electrical influences on the individual cells connected in parallel are difficult to analyze. One reason for this is that cells connected in parallel are usually joined by welding methods. [3] To investigate these influences at cell level, only non-destructive measurement methods are suitable. [4]

II. Methods

The operating mode of batteries and its influences are reflected in the cell aging. [5] Therefore, the aging of a single cell is a good indication of the extent to which an unsymmetrical cell distribution will have a negative impact in a real application. Furthermore the internal resistance is a valuable indicator for the aging state of a Li-ion cell. [6] To determine the internal resistance value of a cell, three different methods (GEIS, AC-IR, DC-IR) are considered in more detail. All three methods are suitable for measuring cell resistances. The test cell is a cylindrical 21700 NMC cell with a capacity of 5 Ah. Before characterizing the test cells, two full cycles are performed with the cells. This initial cycling is used to complete formation processes. [7] Moreover the cell status is evaluated before and after the load profile with a GEIS measurement. The behavior of the cell can be different for the discharge and charge behavior. [8], [9] In this paper, only the discharge behavior of the cell is considered. For all measurements it is important that the open-circuit voltage (OCV) of the specific state of charge (SOC) is reached. This is ensured by a defined relaxation time of the cell after a load. The duration of the relaxation time is based on experience from preliminary tests which is at least one hour. Shorter relaxation times can falsify the measurement results. [10]

Direct Current Internal Resistance (DC-IR) DC-IR or pulse testing is used to determine the internal resistance of Li-ion cells with a short current step. Also, this method is used in the ISO 12405-1 standard to determine the internal resistance of batteries. [11], [12]

The DC-IR method is well suited for modeling long-term effects. [13] For example one long-term effect is the relaxation of cell voltage over time. Out of the voltage drop and relaxation of the cell the characteristics of cells can be determined. This is also displayed in Fig. 1. [14] By using a pulse width of 90 s, the internal resistance and the behavior of the RC element are calculated. For a 2 C discharge pulse with a duration of 90 s, this corresponds to a discharge of 5 % SOC of the cell.

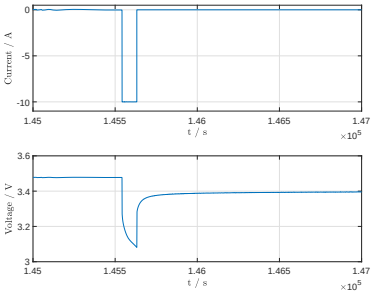


Fig. 1. Voltage response during excitation by a 2C current pulse of a Li-ion cell.

Galvano Electrochemical Impedance Spectroscopy (GEIS) By using the GEIS measurement method, it is possible to characterize Li-ion cells for specific frequencies. Therefore a small sinusoidal AC current is used as the excitation signal of the Li-ion cell. [15] In addition the setting of the GEIS measurement can be seen in Tab.I. [16] During excitation, the system under test responds with a sinusoidal voltage signal shifted by the phase angle ϕ. Based on the current excitation and the voltage response, the behaviour of a Li-ion cell can be determined for the different excitation frequencies. [17] Moreover the excitation currents for the GEIS measurement method should be selected small, since too high excitation currents cause a too large change in concentration of the electrolyte. This then results in inaccuracies of the impedance model. [18] A major advantage of this measurement method is that the cell effects can be measured for very fast load dynamics.

This frequency response of a Li-ion cell is usually shown in a Nyquist diagram. Various physical effects can then be seen, such as mass transport or the frequency behavior of the electrochemical double layer (SEI). All of these effects can be seen in the shape of the Nyquist plot, shown in Fig. 2. [4]

Alternating Current Internal Resistance (AC-IR) The AC-IR method applies a small AC sig-

TABLE I
SETTINGS OF THE GEIS MEASURMENT

Description	Settings
Amplitude	50 mA
Start frequency	10 kHz
End frequency	50 mHz
Points per decade	10
Averaged measurements	5

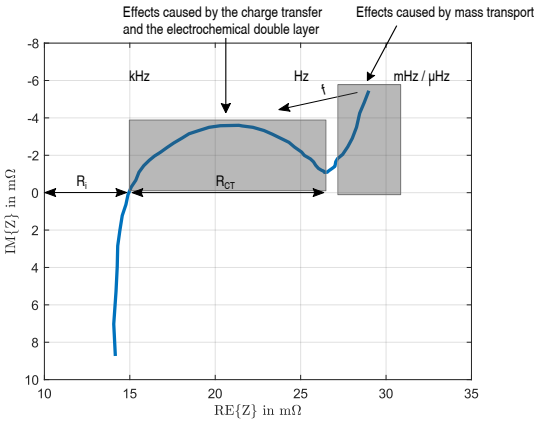


Fig. 2. Nyquist plot and a representation of the physical effects in a Li-ion cell. (according to [4])

nal to the cell or battery system to measure the internal resistance. This method is similar to the GEIS measurement, the main difference being that only one frequency is excited. For the used AC-IR measurement a measuring frequency of 1 kHz is used. The measured resistance R_{AC} at a frequency of 1kHz is almost equal to the pure ohmic resistance R_i for unaged cells. [19] In addition, the pure ohmic resistance is a good indicator of whether the cells are from the same batch. [20] In order to build an accurate simulation model, the response of a single frequency is too small. The advantage of this method is that there are also portable measuring systems on the market that provide a result of the battery resistance in a few seconds. If AC-IR measurement is used to compare internal resistance of cells before and after tests, it is important that the cells have the same SOC before and after the measurement. Furthermore different SOC of cells lead to different internal resistances. [21] All set parameters of the AC-IR measurement are shown in Tab.II.

TABLE II
SETTINGS OF THE AC-IR MEASUREMENT

Description	Settings
Amplitude	max. 160 mA
Measured frequency	1 kHz
Measurement range	30 mΩ
Averaged measurements	5

Based on the Nyquist plot, a simulation model can be created. Thereby, an important indicator for the simulation model is the load type of the cell, as shown in Fig. 3. The load for the simulation is a driving cycle, which lie in the medium frequency range. In other words, the Skin effect behavior at very high frequencies and cell behavior of more than a few hours are not interesting for the simulation of an off-road drivecycle. So a simulation model with two RC element is sufficient to simulate the cell behavior. Furthermore the simulation model is based on a look up table where all parameters are saved. [22]

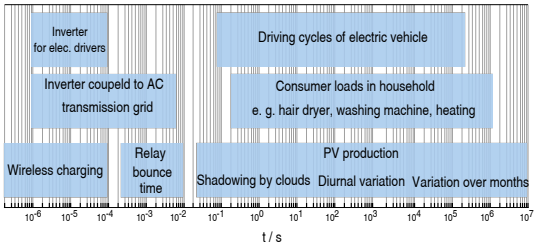


Fig. 3. Typical battery applications, classified according to the dynamics of their electrical workload. (according to [1])

For the design of the test setup, it is important that the electrical load is comparable to the unsymmetrical concept battery. With increasing number of cells

connected in parallel, the load on the individual cells decreases due to Kirchhoff's law. In order to increase the electrical effects of cells with different loads, a 1s3p connection is chosen for the tests. Here, each cell is loaded differently due to the different contact resistances. This described test setup can be also seen in Fig. 5. In addition, the contact resistance value is determined by the length of the leads between the cells. Moreover, the current in battery systems can be several hundred amperes. [23] Therefore, contact resistance is a non-negligible parameter in battery systems. For this reason, the contact resistances in batteries should be as low as possible. To ensure a realistic load on the cells during the test, the distance between two cells in the battery is used to calculate the contact resistance. The material parameters of Hilumin, the cell spacing and material thickness result in a theoretical contact resistance between the cells of $3.6\,m\Omega$. Hilumin is a weldable cell contact material with a low contact resistance that is used in the battery. [24]

The next step is to reduce the environmental influences to ensure that the cell aging is mainly caused by the different cell loading. One goal is to reduce the thermal influences. Also cooling is also provided in real Li-ion batteries. Therefore, all experiments are carried out in a climatic chamber with a constant temperature of 25 °C. In addition, a fan provides simulation of the air-stream and additional cooling. In fact, too high contact resistances lead to a higher power dissipation, which is reflected in a stronger heat development than in a Li-ion battery for automotive applications. Therefore, the contact resistances of the cell holder, which can also be seen in Fig. 4, are investigated. The contact resistances are measured by AC-IR method. Additional, the contact resistance is reduced by using an conductive antioxidant. Besides, a larger wire diameter of $2.5\,mm^2$ is used in the test setup than $0.2\,mm^2$ in the real battery also to reduce the heat generation in the wires.

Another important point is the kind of load during the test. This experimental test setup aims to show the electrical effects of differently loaded cells under

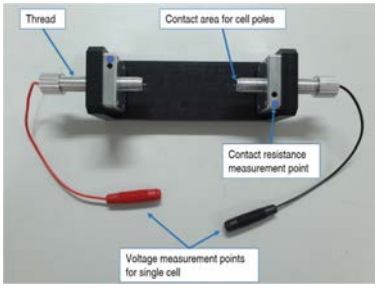


Fig. 4. Selected single cell holder used for parallel connection of cells.

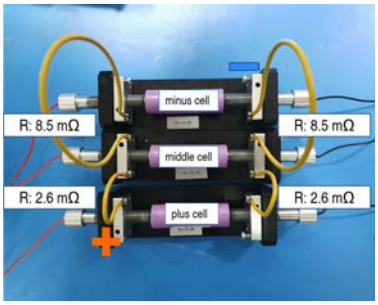


Fig. 5. Defined test setup 02 in 1s3p configuration with different resistances in parallel circuit to simulate the effects of different cell loading in parallel circuit. The resistances are determined by the cable length between the cells and also checked with a micro ohmmeter. Also a temperature sensor Typ K is mounted in the middle of each cell.

real conditions in a parallel circuit. For this reason, a real driving cycle of a motorbike is used. During this load profile, the test setup is discharged for 446 seconds with maximum current pulses of 59.05A. This corresponds to a discharge pulse per cell of approximately 4 C. Furthermore, the load profile is repeated per cycle until 0 % SOC of the test set up is reached. In order to age cells, the driving cycle is repeated 177 times. [25]

III. Results

The primary objective is to show whether unsymmetrically loaded cells also lead to very severe unsymmetrical aging of the cells. This is done by an 1s3p cell configuration where the condition should be similar to a real battery. Therefore the effects of the chosen cell holder (Contact resistance, change in resistance over time and repeatability of contact resistance) are investigated.

A good way to compare the aging state of a cell is the Nyquist plot. Here all changes of the resistance behavior (SEI layer, ohmic internal resistance) are represented over the frequency behavior of the cell. This Nyquist plot can be observed with the GEIS measurement method.

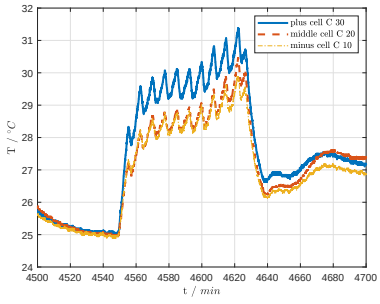


Fig. 6. Temperature behavior of the cells during a complete test cycle with final charging of the cells.

Similarly, cell temperatures are recorded during the test to ensure that unsymmetrical aging is mainly caused by differently loaded cells. The largest deviation between the cell temperatures is shown by the cell at the positive pole with 1.5 °C higher temperature, which is also visible in Fig. 6. Aluminum is used for the anode and cooper for the cathode, since aluminum is a poorer conductor than copper, a higher temperature at the positive pole is to be expected. [26]

Already an unsymmetrical loading, triggered by a deviating cell connector resistance of **1.2 mΩ**, shows effects of different cell aging. This is determined in the first test setup. In the next section, the results of test setup 02 with a deviating cell connector resistance of 5.9 mΩ are evaluated.

As shown in Fig. 7, the capacitive behavior of the cell is lower than at the beginning. Actually, the semicircles should grow with time. It is assumed that after two standard cycles the formation processes of these cells are not yet completed. [27] Therefore, the impedance curves shown are also plausible. For all three cells, the largest change is found in the increase of the pure ohmic response of the cells. In addition, the aging rate for the ohmic response of each test cell is shown in Tab. III. This is also reflected in the power taken from the experimental setup. Thus, a power of 48.80 Wh could be drawn at the beginning and only 33.74 Wh after 177 cycles. Moreover, this corresponds to a power loss of more than 30 %. With a power loss of 30 %, the End of Life (EOL) criteria of the test setup is also fulfilled. Based on the result of this tests, too large deviations in the contact resistances lead to unsymmetrical cell aging.

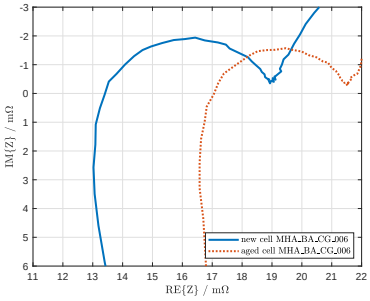


Fig. 7. Analysis of the aged cell based on the GEIS measurement data. These measurement data are based on a SOC of the cell of 60 % after 177 cycles.

For the **simulation part**, a characterization of the cell is required. For this purpose, two methods are used and compared: the GEIS measurement and the DC-IR technique. Their simulation results are shown in Fig. 8. There, a section of the selected driving

TABLE III
EVALUATION OF THE TEST RESULTS FOR TEST SETUP 02.

Description	Value
Name	MHA_BA_CG011
Position in test setup	minus
Start resistance	14.69 mΩ
Increase of resistance after 177 cycle	31.08 %
Name	MHA_BA_CG006
Position in test setup	middle
Start resistance	14.46 mΩ
Increase of resistance after 177 cycle	27.10 %
Name	MHA_BA_CG009
Position in test setup	plus
Start resistance	14.74 mΩ
Increase of resistance after 177 cycle	18.66 %

cycle with a three second pause can be seen. It can be seen that the simulated GEIS data doesn´t reflect the actual relaxation process of the cell. This is in contrast to the simulation with the DC-IR method, where the relaxation process during the pause is better represented.

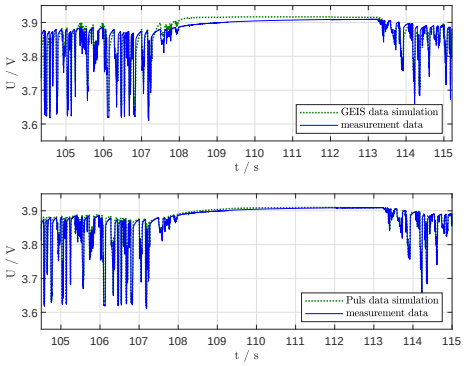


Fig. 8. Comparison between the simulated data and the real measurement

Two methods (DC-IR and EIS) are used to obtain the parameters for the simulation model. The simulation model is based on RC elements. In order to obtain the cell properties for the simulation model, it is necessary to check them at different cell SOCs. Since the cell behavior changes over the SOC. This can also be seen in the Fig. 9. In addition the cell characterization procedure is described in Tab.IV. Also the voltage profile for the cell characterization is shown in Fig. 10. At the end of each discharge step a GEIS measurement is done followed by the discharge pulse.

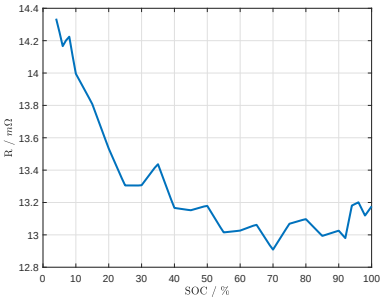


Fig. 9. Real resistance R_E behavior of a cylindrical 21700 NMC Li-ion cell over the SOC at 1 kHz.

TABLE IV
DETERMINATION OF CELL PARAMETERS

Step	Description	Parameter and Conditions
1	full charge	@0.5C, CC+CV
2	2 standard cycle	@1C, CC+CV
3	rest	15 min
4-8	dch to 90 % SOC	@2 % SOC + 2.5 h pause
9-25	dch to 10 % SOC	@5 % SOC + 1 h pause
26-30	dch to 0 % SOC	@5 % SOC + 2.5 h pause
30	ch to 30 % SOC	@1C, CC+CV

It is shown that a simulation for the driving cycle both methods are suitable to obtain the parameters of the RC elements. The simulation results show that the DC-IR method reflects the long-term behavior

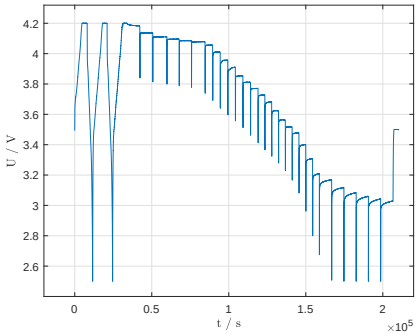


Fig. 10. Cycle for characterization of lithium ion cells

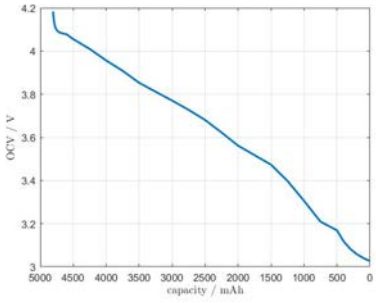


Fig. 11. Open Circuit Voltage over the capacity of the Li-ion cell

of cells better than the simulation with the GEIS parameters.

IV. Conclusion

These findings lead to the conclusion that aging due to differentially loaded cells in parallel connection leads to accelerated and unsymmetrical cell aging. In addition, similar tests should be performed during charging operations. A slightly different heating of the individual cells is also observed; it must be identified whether this effect is amplified in a real battery system.

In my opinion, the advantages are outweighed by the 30 % higher capacity of the unsymmetrical battery compared to the symmetrical battery design. Of course, further research is needed on how to define the voltage limits of the individual battery modules and whether a lower EOL criterion should be applied in order to operate the unsymmetrical battery safely.

In developing the RC element model, one also encounters other mathematical descriptions that reflect cell behavior. One possibility for the cell simulation model is the constant phase element. Here it is interesting to see how it compares to the RC element in terms of simulation time and simulation accuracy.

In summary, the tests with the minimal test setup cannot replace any test standards for a Li-ion battery. Rather, the test results serve to map the environmental impact of the system over the course of the product's life and to identify any weak points already in the development phase.

Acknowledgment

The author would like to thank the KTM Forschungs & Entwicklungs GmbH.

References

[1] M. J. Brand, "Lithium-ion battery cells and systems under dynamic electric loads," Ph.D. dissertation, Technische Universität München, 2018.

[2] X. Gong, R. Xiong, and C. C. Mi, "Study of the characteristics of battery packs in electric vehicles with parallel-connected lithium-ion battery cells," in *2014 IEEE Applied Power Electronics Conference and Exposition - APEC 2014*. IEEE, mar 2014.

[3] P. A. Schmidt, "Laserstrahlschweißen elektrischer Kontakte von LithiumIonen-Batterien in Elektro- und Hybridfahrzeugen," Ph.D. dissertation, Technische Universität München, 2015.

[4] A. Jossen, "Fundamentals of battery dynamics," *Journal of Power Sources*, vol. 154, no. 2, pp. 530–538, mar 2006.

[5] C. R. Birkl, M. R. Roberts, E. McTurk, P. G. Bruce, and D. A. Howey, "Degradation diagnostics for lithium ion cells," *Journal of Power Sources*, vol. 341, pp. 373–386, feb 2017.

[6] P. Keil, "Aging of lithium-ion batteries in electric vehicles," Ph.D. dissertation, TU München, 2017.

[7] E. Peled and S. Menkin, "Review—SEI: Past, present and future," *Journal of The Electrochemical Society*, vol. 164, no. 7, pp. A1703–A1719, 2017.
[8] R. Korthauer, Ed., *Handbuch Lithium-Ionen-Batterien*. Springer Berlin Heidelberg, 2013.
[9] P. Keil, "Aufbau und parametrierung von batteriemodellen," *Technische Universität München*.
[10] F. M. Kindermann, A. Noel, S. V. Erhard, and A. Jossen, "Long-term equalization effects in li-ion batteries due to local state of charge inhomogeneities and their impact on impedance measurements," *Electrochimica Acta*, vol. 185, pp. 107–116, dec 2015.
[11] "DIN ISO 12405-1:2011-08 Elektrische Straßenfahrzuege - Prüfspezifikation für Lithium-Ionen Antriebsbatteriesystem und Batterieteilsysteme - Teil 1: Hochleistungssysteme."
[12] A. Lasia, *Electrochemical Impedance Spectroscopy and its Applications*. Springer New York, 2014, p. 333 - 350.
[13] D. Kehl, T. Jennert, F. Lienesch, and M. Kurrat, "Electrical characterization of li-ion battery modules for second-life applications," *Batteries*, vol. 7, no. 2, p. 32, may 2021.
[14] J. P. Schmidt, "Verfahren zur Charakterisierung und Modellierung von Lithium-Ionen Zellen," 2013.
[15] S. Wang, J. Zhang, O. Gharbi, V. Vivier, M. Gao, and M. E. Orazem, "Electrochemical impedance spectroscopy," *Nature Reviews Methods Primers*, vol. 1, no. 1, jun 2021.
[16] S. Gantenbein, "Impedanzbasierte Modellierung von Lithium-Ionen Zellen und deren Degradationsverhalten," 2019.
[17] N. Sassano, "Entwicklung eines Messsystems zur funksynchronisierten elektrochemischen Impedanzspektroskopie an Batterie-Zellen," Master's thesis, Fachhochschule Westküste, 2015.
[18] J. P. Schmidt, "Verfahren zur Charakterisierung und Modellierung von Lithium-Ionen Zellen," Ph.D. dissertation, Karlsruher Instituts für Technologie (KIT), 2013.
[19] *Batterie-Messungenin der Praxis*, BRS Messtechnik GmbH, Nov. 2015.
[20] B. S. Paul, "Analyse der Asufallwahrscheinlichkeiten von Lithium-Ionen-Energiespeichern in elektrifizierten Fahrzeugen," Ph.D. dissertation, Universität Ulm, 2014.
[21] *From 1 hour to Just 10 Seconds: Using the Low-frequency AC-IR Method as aQuicker and More Stable Alternative to DC-IR Testing of Lithium Ion Batteries*, HIOKI.
[22] H. Witzenhausen, "Elektrische Batteriespeichermodelle: Modellbildung, Parameteridentifikation und Modellreduktion," Ph.D. dissertation, RWTH AACHEN UNIVERSITY, 2017.
[23] P. D.-I. O. Bohlen, "Skript zur Lehrveranstaltung: Energiespeicher," 2019.
[24] T. STEEL, *Hilumin The best option for battery applications and connector strips*.
[25] T. L. M. Gewald, "Beschleunigte Alterungscharakterisierung von Lithium-Ionen-Zellen in der Fahrzeuganwendung," Ph.D. dissertation, TU München, 2021.
[26] S. Leuthner, "Übersicht zu Lithium-Ionen-Batterien," in *Handbuch Lithium-Ionen-Batterien*. Springer Berlin Heidelberg, 2013, pp. 13–19.
[27] T. Li, X.-Z. Yuan, L. Zhang, D. Song, K. Shi, and C. Bock, "Degradation mechanisms and mitigation strategies of nickel-rich NMC-based lithium-ion batteries," *Electrochemical Energy Reviews*, vol. 3, no. 1, pp. 43–80, oct 2019.

TABLE V
LIST OF SYMBOLS

Symbol	Name	Unit
R_i	series internal cell resistance	Ω
R_{CT}	Resistance caused by the SEI layer	Ω
t_{pulse}	Discharge pulse duration	s
t_{fit}	Duration for the characterization of RC elements	s
I_{pulse}	Discharge pulse	A
R_{AC}	Measured resistance at 1 kHz	Ω
R_E	Real resistance of the cell	Ω

Markus Hausladen is student of the Department of Mechatronics at MCI Innsbruck/Austria. He is currently writing his master thesis at KTM Forschungs & Entwicklungs GmbH.

Concept Design and Evaluation of isolated DC/DC Converters in mid power DC charging stations

Andreas Hofer, Maurizio Incurvati (supervisor), Tizian Senoner (supervisor)

***Abstract*—Mid power DC charging stations are recently gaining popularity, as high power DC charging is not suitable for every customer. High power DC charging is associated with high cost due to the needed equipment and infrastructure. Thus, a trend towards mid power charging stations with charging power up to 50 kW can be seen. As the market is still evolving and not many products can be found in this segment yet, general knowledge of how to implement such charging stations is still not established. Therefore, DC/DC converters which are to be employed in such systems are analyzed and compared to gain deeper knowledge. An LLC and a dual active bridge (DAB) converter are compared by modelling and design calculations. Basic topology simulations are performed in LTSpice to validate the design concept and examine qualitative properties of waveforms on key components. Ultimately, measurements on real hardware are performed in order to determine their advantages and disadvantages in the context of mid power DC charging. The gathered data is used to suggest improvements on examined hardware and ultimately decide which topology is better suited for the application. Reasoning for opting for the LLC topology is provided by means of power density, efficiency and electromagnetic interference (EMI) evaluation in comparison to the analyzed DAB stage.**

***Index Terms*—Mid Power DC Charging, LLC, DAB, Dual Active Bridge, Resonant Converter Design, Electric Vehicle, Electric Vehicle Supply Equipment**

ABBREVIATIONS

AC-charging	convetional single or three phase AC charging
DAB	dual active bridge
DUT	device under test
EMC	electro-magnetic compatibility
EMI	electro-magnetic interference
e-mobility	electromobility
EV	battery electric vehicle
EVs	battery electric vehicles
HPDC	high power DC fast charging
MPDC	mid power DC charging
SiC	silicon carbide

I. INTRODUCTION

RECENT developments in electromobility (e-mobility) show a drift of charging power levels [1].

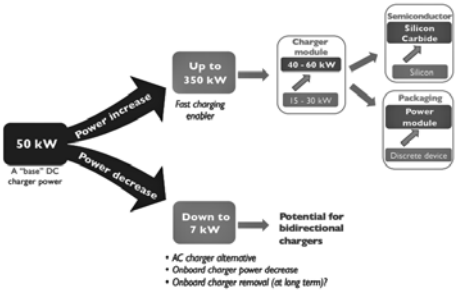


Figure 1. Trends in DC EV Charging Power [1]

Extremely high charging power of up to 350 kW is seeked for in fast battery electric vehicle (EV) refueling in high power DC fast charging (HPDC) applications. A modular converter design approach is implemented in order to facilitate modularity and scalability of chargers and charging powers. Discrete,

silicon based semiconductor switches are ousted by wide band gap technology switching modules, such as silicon carbide (SiC) modules [2]. On the other hand, a trend towards charging power levels below $50\,\mathrm{kW}$ is visible in mid power DC charging (MPDC) applications for battery electric vehicles (EVs) [1]. Reasoning can be found in the elevated cost of HPDC units and their corresponding infrastructure. On-board AC chargers are becoming less powerful and might be obsolete in future due to developments in MPDC [1]. In contrast to HPDC, MPDC devices are not readily available as the market is still evolving. A general understanding of optimum charger implementation is yet to be established. HPDC is tailored for fast EV refueling in charging parks or at gas stations, where charging speed is of the essence. High cost of equipment and infrastructure is secondary due to the high utilization and on-charging of the elevated costs to customers. convetional single or three phase AC charging (AC-charging) is more suitable for charging at residence, where individuals spend a significant amount of time and charging can take several hours, even a whole night. MPDC is oriented towards business locations and shopping centers, where customers on average will spend up to 3 hours [3], therefore tolerating the increased charging times which arise compared to HPDC. Furthermore, it features a practical alternative for accommodation businesses which intend to offer faster EV refueling to their customers. Cost of infrastructure, although being comparably high in comparison to AC-charging, is secondary as a high power grid connection along with according cabling might already be present in many business due to elevated energy demand of other terminal devices. Currently, only a very limited amount of MPDC chargers is available for retail. Data on implemented DC/DC converter topologies is scarce. Recent studies [1] [2] show an equal market share of both the LLC resonant converter topology and the dual active bridge (DAB) converter topology in MPDC. Both the DAB and the LLC topology offer promising features for MPDC applications, further elaborated in the course of this thesis. Hard switching topologies are usurped by resonant and soft switching topologies due to their higher efficiency and higher power density along with lower electro-magnetic interference (EMI) [2].

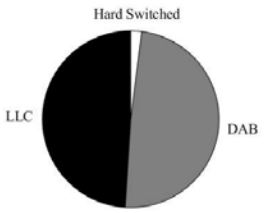


Figure 2. Prevailing DC/DC Topologies in MPDC [1] [2]

Gaining deeper knowledge in the efficient design and implementation of such chargers could pose future advantages in terms of competitiveness in the segment. An early entry into this evolving market might grant a significant head start. Knowledge gained in MPDC development might be transferable to other energy conversion appliances.

II. Methods

Existing MPDC are evaluated according to their charging characteristics and employed DC/DC converter topologies. An LLC and a DAB stage MPDC are selected and subjected to a number of tests. First, key operating points related to MPDC are defined based on data from over $3.9\,\mathrm{Million}$ real world charging sessions [4]. The design process of both topologies is evaluated. Basic topology simulations are performed to validate the proposed design. Waveforms on key components in the DC/DC stage of both device under test (DUT) are captured along with their efficiency, EMI spectrum and power density. Optimization suggestions for the examined hardware are elaborated based on gathered data. For converter design, lossless models for both the DAB

and the LLC stage are presented. The model used for observations in the LLC converter design is shown below.

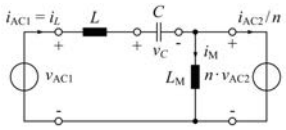


Figure 3. Lossless LLC model [5]

After determining fundamental equations for LLC design, losses in switching components are evaluated and key component values are optimized following [6].

Similarly, the DAB is analyzed using the lossless model shown in Fig. 4.

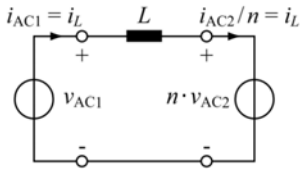


Figure 4. Lossless DAB model [5]

The resulting equations are again used in combination with loss calculations for designing a DC/DC stage featuring a maximum operating area at minimum losses in components.

Both designs are simulated in LTSpice for preliminary validation and further theoretical understanding of converter behaviour.

The simulation setup for the examined LLC stage is shown in Fig. 5.

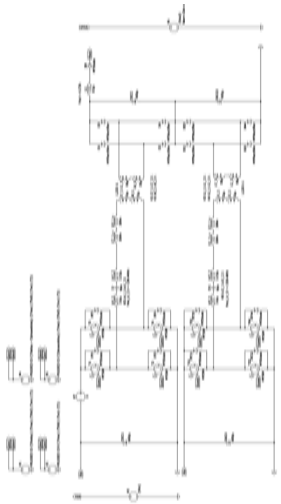


Figure 5. LLC Simulation Setup

The simulation setup for the examined DAB stage is shown in Fig. 6.

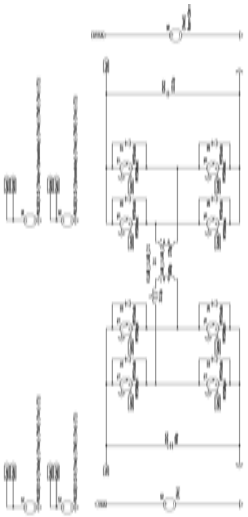


Figure 6. DAB Simulation Setup

The measurement setup used during evaluations is shown in Fig. 7.

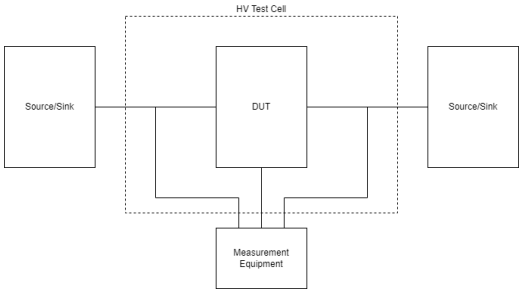


Figure 7. Measurement Setup

Measurements and simulations mentioned above are established at output voltages of 333 V, 420 V and 900 V, all commonly used charging voltages in field [4]. Each measurement and simulation is taken at maximum output power of the DUT. For presentation of results in this document, 420 V output voltage is used as it is currently the most used in field [4]. Waveforms and other data captured in simulations and measurements on real hardware is evaluated and used to decide on an optimum topology for MPDC DC/DC converters. Furthermore, it is used to provide optimization suggestions for the implemented converters. Results of previous work [5] [7] and similar studies [8] are considered.

III. Results

Waveforms simulated for the LLC stage at an output voltage of 420 V are shown in Fig. 8.

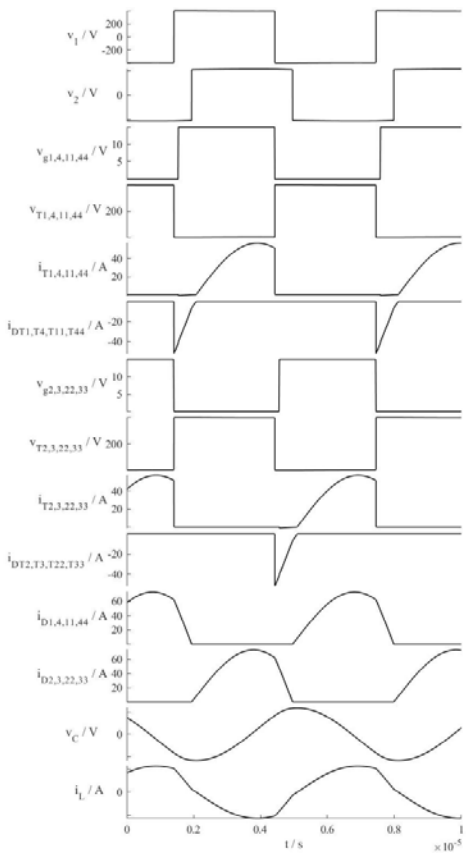


Figure 8. LLC: Simulation results at 420 V output voltage

Simulated LLC waveforms show soft turn-on of all switches along with moderate current gradients due to the resonant nature of the LLC stage. Expected EMI is low due to overall smooth transitions in switching operations. The measured counterparts are shown in Fig. 9

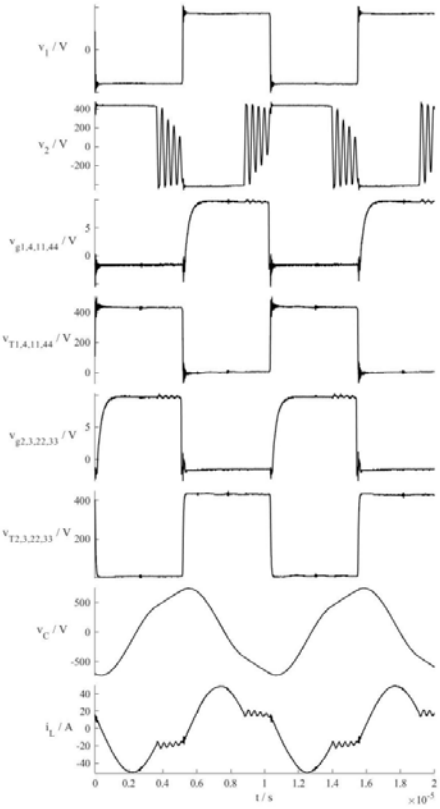


Figure 9. LLC: Measurement results at 420 V output voltage

In comparison to the previously simulated results, the secondary transformer voltage v_2 and the resonant inductor current i_L of the LLC stage feature substantial ringing, probably caused by parasitic effects. Soft switching is also observed in measured results in form of voltages across switches falling to nearly zero before switching operations. The frequency of the observed ringing lies in the high kHz to low MHz range and is expected to appear in the EMI spectrum at DC output terminals. The captured conducted EMI spectrum at the DC output is shown in Fig. 10

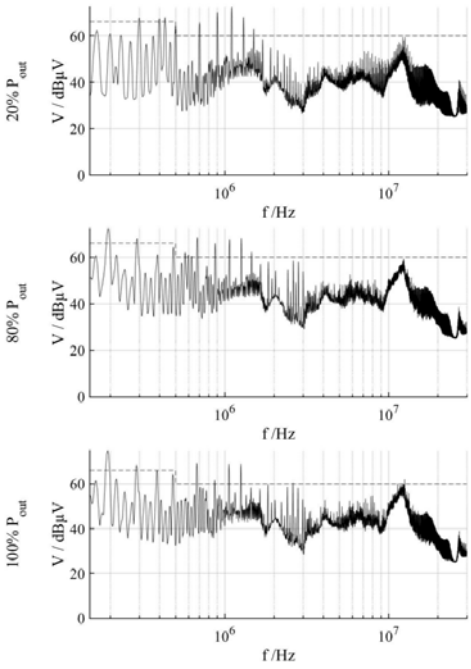


Figure 10. LLC: Conducted EMI at 420 V output voltage

As required by European standard [9], the conducted EMI spectrum is evaluated at 20 % and 80 % output power. Although being expected to feature decent electro-magnetic compatibility (EMC) performance, the LLC stage is still in need of EMI filtering in order to achieve EMC. Due to the absence of EMI filtering in the examined LLC stage, it is non-compliant to EMC regulations [9]. Exceeding of the defined limits in the lower kHz range is caused by lower switching frequency harmonics, with both odd and even harmonics of the switching frequency appearing in the spectrum. Peaks of the odd switching frequency harmonics are however more pronounced. The non-compliance in the upper kHz and lower MHz range is likely caused by appearance of the parasitic ringing as stated above.

Waveforms simulated for the DAB stage at an output voltage of 420 V are shown in Fig. 11.

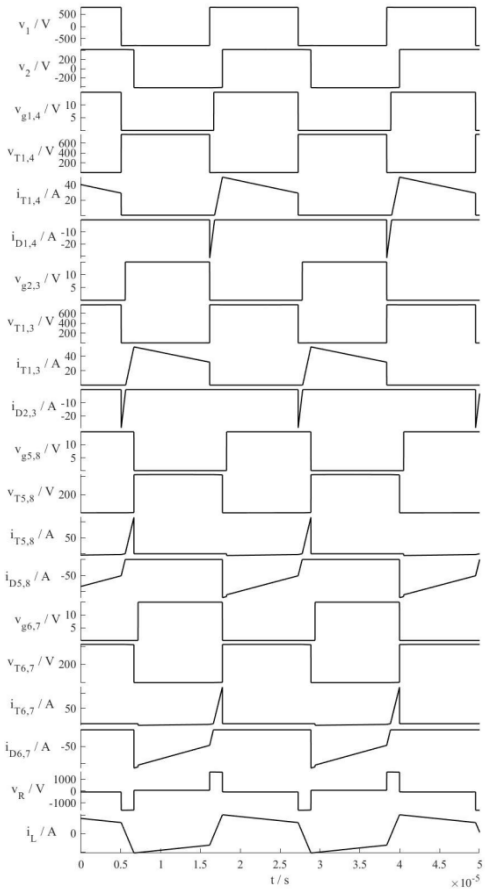


Figure 11. DAB: Simulation results at 420 V output voltage

When compared to the examined LLC stage, current gradients in the DAB are drastically increased, also visible in measured and simulated results, where current transitions in the LLC resemble sinusoidal shapes while DAB currents feature steep trapezoidal transitions. Although soft-switching is again achieved, the expected EMI performance is inferior to the LLC stage. This is due to faster transients and increased gradients in observed currents and voltages.

The measured counterparts are shown in Fig. 12

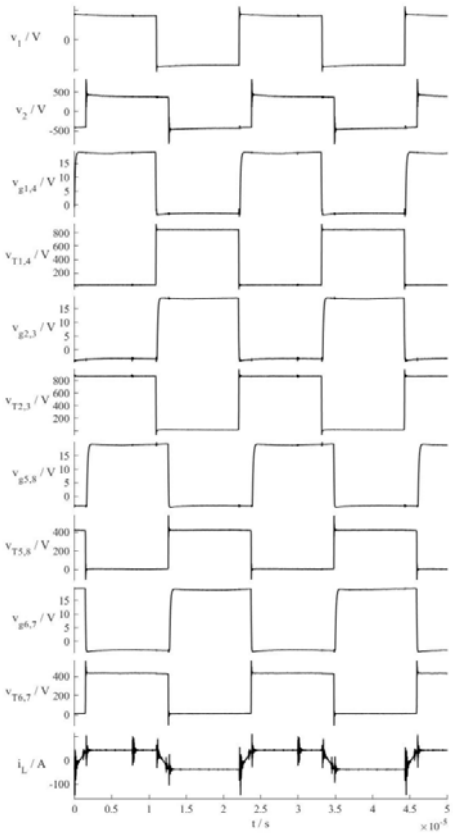


Figure 12. DAB: Measurement results at 420 V output voltage

The steep transients in switch-voltages mentioned above most likely cause the slight overshoot at switch turn-on and turn-off in the observed DAB stage. As in the LLC stage, parasitic impedances along with the fast transitions in captured waveforms cause ringing of the inductor current i_{L}. The captured conducted EMI spectrum at the DC output is shown in Fig. 13

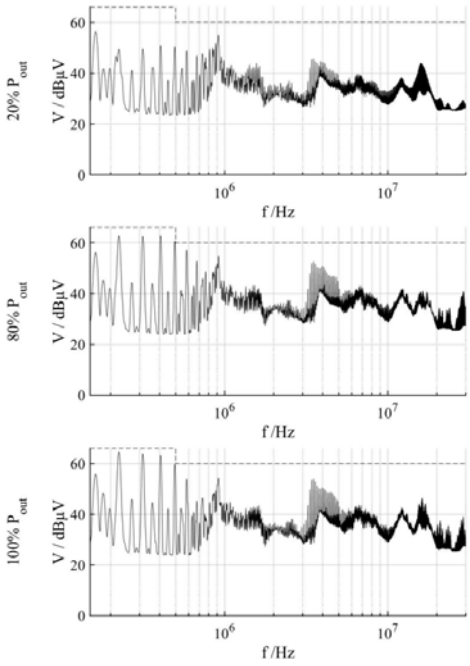


Figure 13. DAB: Conducted EMI at 420 V output voltage

Even though the DAB stage is expected to have inferior EMI performance due to its high current gradients and fast switching transitions, it achieves EMC at 20 % and 80 % output power. This is due to the extensive EMI filtering implemented withing the DUT. The power density of the examined LLC stage is nearly 4 times higher than the power density of the examined DAB stage. Not only is this related to topology characteristics, but also to the switching frequency, which is up to 4 times higher in the implemented LLC. Key characteristics of the two examined DC/DC stages are evaluated. Their performance is assessed based on soft switching performance, component stresses, current gradients, expected EMI and control complexity, as-well as power density and efficiency.

Key evaluated criteria are summarized in Fig. 14.

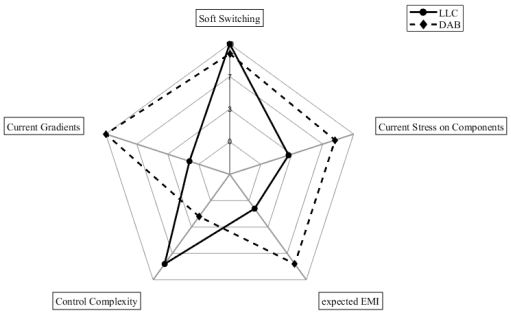


Figure 14. Radar Plot of evaluated Characteristics

IV. Conclusion

Based on results in Sec. III, the LLC stage is deemed more promising for MPDC implementation. Although measured EMI is higher in the examined LLC stage, EMI performance is still expected to be superior due to the absence of EMI filtering in the examined LLC stage, while the examined DAB features extensive EMI filtering. By employing wide band gap technology switches such as SiC, the performance of the LLC stage could be further enhanced. Due to the paralleling of several phases in the examined LLC converter, deviations in resonant tank frequencies might lead to major challenges in system control. Optimization of the examined DAB stage by increasing switching frequency is possible only to a certain extent, as optimization simulations [10] show a peak performance of the DAB at the chosen switching frequency. Furthermore, the switching frequency was chosen to avoid the appearance of its low harmonics in the conducted EMI spectrum and thus further challenges in achieving EMC. Even though the bidirectionality of the DAB stage might be advantageous in some applications, it is not expected to be commonly used in MPDC. Furthermore, bidirectional energy transfer can also be achieved by modifying LLC stage [11], also increasing its performance due to synchronous rectification [12]. As a follow-up of the performed investigations, a SiC based LLC converter will be developed and tested in

order to for it to be considered as the optimal DC/DC converter stage for future MPDC related projects.

ACKNOWLEDGMENT

The author would like to thank supervisors M.Incurvati and T.Senoner for their professional support during the development of this document. Special regard is also payed to A. Oberrauch and the alpitronic team, particularly to P. Niedermayr, P. Costanzi, A. Lösch, H. Kornprobst and S. Dissertori Sölva for their engagement and endorsement during work at alpitronic GmbH.

REFERENCES

[1] A.Ly, M.Rosina, "DC CHARGING FOR PLUG-IN ELECTRIC VEHICLES 2021," 2021.
[2] M. Brenna, F. Foiadelli, C. Leone, M. Longo, "Electric Vehicles Charging Technology Review and Optimal Size Estimation," 2020.
[3] Statista Research Department, "Average time spent at regional and super-regional malls by consumers in the United States from April to June in 2016, by age," 2017.
[4] alpitronic GmbH, "Sample Data from 3.9 million charging sessions, hyperlog Management System," 2022.
[5] F. Krismer, "Modeling and optimization of bidirectional dual active bridge DCDC converter topologies," 2010.
[6] S. Abdel-Rahman, "Resonant LLC Converter: Operation and Design 250W 33Vin 400Vout Design Example," 2012.
[7] P.Niedermayr, "Modelling of the Double Active Bridge for HYC_50," 2020.
[8] S.Ditze, "Studie "HyperCharger"," 2017.
[9] VDE, "DIN EN 61000-1-2 VDE 0839-1-2:2017-07 Elektromagnetische Verträglichkeit (EMV)," 2017.
[10] P.Niedermayr, "HYC_50_power_loss_calculations," 2020.
[11] J.Jung, H.Kim, J. Kim,M. Ryu, J. Baek, "High Efficiency Bidirectional LLC Resonant Converter for 380V DC Power Distribution System Using Digital Control Scheme," 2012.
[12] Y.Liu, "High Efficiency Optimization of LLC Resonant Converter for Wide Load Range ," 2007.

Andreas Hofer, BSc. is with the R&D Department of alpitronic GmbH Bozen/Italy. Among others he is responsible for activities in hardware and system design and testing. He is currently working on his Master of Science in Engineering at MCI/Innsbruck where he also received his Bachelor's degree.

Entwicklung einer Ausbildungsplattform für parallele Industrieroboterkinematiken

Serafin Kollegger, Benjamin Massow (supervisor)

***Kurzfassung*—Für die Automatisierung von industriellen Produktionen ist der gezielte Einsatz von Robotersystemen wichtig. Auf Grund der Anzahl an unterschiedlichen Anforderungen existieren viele Bauweisen von Manipulatoren, und Steuerungsstrategien. Um eine optimale Automatisierungslösung für spezifische Anwendungen entwickeln zu können, ist ein vertiefendes Know-How über Roboterkinematiken und Steuerungsstrategien notwendig. Die in dieser Arbeit entwickelte Ausbildungsplattform stellt die Möglichkeit zur Verfügung, vertiefende Einblicke in die Auslegung und Umsetzung von Parallelkinematik-Robotern für industrielle Anwendungen zu erlangen. Die Ausbildungsplattform wird durch die Kombination aus Ausbildungskonzept und zugehöriger Hard- und Software entwickelt. Die Steuerung und Regelung des Manipulators wird durch eine Modell basierte Entwicklungsstrategie mit Hilfe der Entwicklungs- und Simulationssoftware MATLAB/*Simulink*, und der integrierten Toolbox *Simscape* entwickelt. Durch ein Interface zum Datenaustausch zwischen *Simulink* und *TwinCAT* können Anwender eigenentwickelte Funktionen und Regler ohne Schwierigkeiten implementieren. Dafür wird eine Steuerung für den Roboters entwickelt, wodurch die gesamte Programmstruktur ohne Barriere zugänglich ist. Durch ein Simulationsmodell ist es für Anwender möglich die eigenentwickelten Funktionen in einer sicheren Simulationsumgebung zu entwerfen und zu testen. Die Ausbildungsplattform inklusive Simulationsumgebung wurde anhand einer Ausbildungsaufgabenstellung getestet, und konnte durch die einfache Anwendung und Umsetzung überzeugen.**

***Schlagwörter*—industrielle Automatisierung, Parallelroboter, Ausbildungsplattform, Modell basierte Entwicklung**

S. Kollegger studiert am MCI Management Center Innsbruck, e-mail: se.kollegger@mci4me.at

B. Massow ist Hochschullektor am Studiengang Mechatronik, MCI, Innsbruck, Austria.

I. Einleitung

DIE Notwendigkeit der Automatisierung für produzierende Unternehmen im europäischen Raum ist eine große Herausforderung. Zum einen gibt es für viele industrielle Güter starke Konkurrenz aus nicht europäischen Ländern, und andererseits herrscht ein großer Mangel an Facharbeitskräften in Europa. Diese Kombination hat zur Folge, dass immer mehr Unternehmen verschiedener Größen und Branchen ihre Produktion, durch den Einsatz von automatisierten Anlagen, effizienter und effektiver machen müssen.

Dabei spielen Roboter eine zentrale Rolle in der Prozesskette von automatisierten Produktionsanlagen durch Werkstückmanipulation, Montage zweier Bauteile, Bearbeitung des

Werkstücks, und ähnliches. Diese unterschiedlichen Aufgaben von Roboter erzeugen komplexe und häufig gegensätzliche Anforderungen an die Mechanik, Kinematik, und Steuerungssysteme. Somit ist eine gezielte Entwicklung und Auswahl von Robotersystemen erforderlich, um eine optimale Automatisierungslösung zu generieren. Der Entwicklungsprozess erfordert eine ganzheitliche Analyse der Prozessanforderungen, wofür der Produktionsprozess, sowie die Automatisierungsmöglichkeiten verstanden werden müssen. Um letzteres Wissen und die erforderlichen Fähigkeiten zur Umsetzung erlangen zu können, ist eine praxisnahe Ausbildung unumgänglich.

A. Motivation

Im Bereich der Parallelkinematik-Roboter wurden in den letzten Jahren viele neue Konzepte entwickelt, welche bisher vielversprechende Ergebnisse liefern. Somit ist es möglich, viele Anwendungen welche bislang nur schwer mit herkömmlichen seriellen Industrieroboter umzusetzen sind, durch Parallelkinematik-Roboter zu ersetzen. Dadurch kann sowohl eine höhere Effektivität erreicht werden, als auch Investitionskosten gesenkt werden [1]. Jedoch befinden sich viele der vorgeschlagenen Konzepte noch im Prototypenstadium und sind noch nicht industriell einsetzbar. Dabei besteht eine große Wissenslücke in der Entwicklung von idealen Parallelkinematik-Konfigurationen im Zusammenspiel mit Lagerarten und deren Materialkompositionen für spezifische Anforderungen. Zusätzlich besteht Forschungsbedarf in der Modellierung für starre und elastische Modellsimulationen, um dynamischen Effekte besser verstehen zu können und somit die Entwicklung neuer Kinematik-Bauweisen zu ermöglichen. Des Weiteren gibt es bei der Integration bestehender industrieller Parallel-Roboter vergleichsweise wenig Dokumentationen und Fallstudien. Weshalb viele Betriebe die Möglichkeiten der Automatisierung mittels dieser Roboter nicht einsetzen können.

Mit Hilfe der in dieser Arbeit entwickelten Ausbildungsplattform soll es möglich sein eine ganzheitliche Betrachtung für Robotersysteme mit paralleler Kinematik zu schaffen. Dazu wird eine Ausbildungsaufgabe formuliert, anhand welcher die Auszubildenden einen strukturierten Entwicklungsprozess von Robotersystemen durchführen. Um dies realisieren zu können, soll ein virtuelles Modell des Manipulators erzeugt werden. Die Steuerungsfunktionen können während der Entwicklung am Modell validiert und getestet werden. Zusätzlich soll den Auszubildenden ein Analysewerkzeug für parallele Mechanismen zur Verfügung gestellt werden, womit die Anforderungen an die zu entwickelnde Bewegungssteuerung ermittelt werden können. Durch die Kombination aus virtuellen und realen Komponenten soll es möglich sein, die Ausbildungsplattform für *Remote Laboratories* verwenden zu können, ohne dass der Praxisbezug verloren geht.

Damit dieses Ziel erreicht werden kann, ist die Ausbildungsplattform nicht nur eine physische Anlage, sondern umfasst die zusätzlichen Komponenten,

- ein virtuelles Modell,
- eine Softwareentwicklungsumgebung und
- eine Softwaresimulationsumgebung.

Durch diesen Ansatz der modellbasierten Entwicklung ist es für einen Auszubildenden möglich, durch geführte Eigenentwicklungen von Modellen, Steuerungen und Regelungen, eine Roboterbewegungssteuerung selbstständig zu

entwickeln und an der realen Anlage durch eine *Pick and Place* (P&P) Anwendung zu testen. In Abbildung 1 ist das Konzept der Plattform schematisch dargestellt. In [2] wurde bereits ein ähnlicher Ansatz umgesetzt.

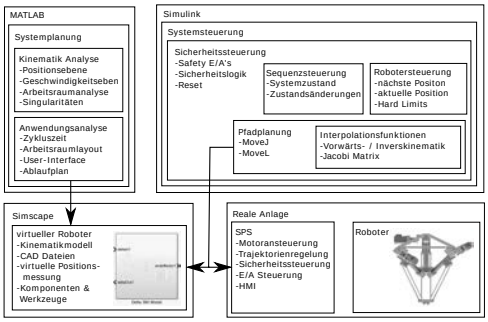


Abbildung 1. Konzeptdarstellung der Ausbildungsplattform.

II. Grundlagen und Anforderungen

Um effektiv und effizient automatisierte Anlagen zu entwickeln, ist der Ansatz einer modellbasierten Systementwicklung laut [3] mit vielen Vorteilen verbunden. Einerseits gibt es die Möglichkeit, Modelle, Software und Steuerungshardware testen zu können bevor diese in der Anlage integriert sind. Zum anderen ist für die Modellierung eine detaillierte und ganzheitliche Betrachtung der Komponente erforderlich. Durch diese können bereits zu Beginn des Entwicklungsprozesses Abweichungen der Anforderungen entdeckt und korrigiert werden. Eine mögliche Vorgehensweise der modellbasierten Systementwicklung ist von [2] aufgestellt worden. In dieser müssen fünf Phasen durchlaufen werden, beginnend mit der ersten Phase der Definition von anwendungsabhängigen Spezifikationen. In Phase Zwei wird das System modelliert, und in eine Simulationssoftware eingebunden. Die dritte Phase beschäftigt sich mit der Auswahl der Aktuatoren und Sensoren. In der vierten Phase werden Regelungen für das System entwickelt und zuletzt behandelt die fünfte Phase die Simulation des gesamten Systems mit Verifizierung durch Prototypen und reale Komponenten.

A. Kinematikanalyse

Die Kinematikberechnungen werden benötigt, um die Anwendung analysieren zu können und prozessrelevante Anforderung mit quantitativen Werten zu beschreiben. Erster Schritt der Kinematikanalyse ist die Mobilitätsanalyse. Hier werden Eigenschaften eines Delta 360 Manipulators durch wenige Kennzahlen beschrieben. Die Dimensionen des Roboters können der Tabelle I entnommen werden.

Tabelle I
Dimensionen des Manipulators.

Komponente	Zeichen	Wert	Einheit
virtuelle feste Plattform	r_b	222.106	mm
virtuelle bewegliche Plattform	r_m	42	mm
Stäbe	l	400	mm
Antriebsreichweite	r	150	mm
Antriebsanordnung	θ	120	°
Anstellwinkel	ϕ	45	°

Die Kennzahlen der Mobilitätsanalyse werden später zum Aufstellen des Gleichungssystems zur Berechnung der Kinematikfunktionen benötigt. Die Ermittlung der Freiheitsgrade und somit die Anzahl der angetriebenen Achsen kann durch die Kutzbach-Grübler Gleichung für räumliche Mechanismen aus [4] mit

$$m = \lambda n_{mb} - (5j_1 + 4j_2 + 3j_3)) \qquad (1)$$

durchgeführt werden, wobei λ die Anzahl der möglichen Freiheitsgrade des Arbeitsraumes sind, d.h. für räumliche Manipulatoren sechs. Die Zahl n_{mb} beschreibt die Anzahl der Beweglichen Bauteile, welche in diesem Fall sieben sind, und j_i beschreibt die Anzahl an Freiheitsgrade der Gelenke mit i Freiheitsgraden. Somit kommt der Delta 360 Manipulator auf drei Freiheitsgrade und benötigt drei angetriebene Achsen. Mit der Summe der Freiheitsgrade j_{tot} und der Anzahl der beweglichen Bauteile kann die Anzahl n_{ikl} der unabhängigen Kinematikschleifen (engl. indepentent kinematic loops)(IKL) durch

$$n_{ikl} = j_{tot} - n_{mb} = 9 - 7 = 2 \qquad (2)$$

bestimmt werden. Danach können die Zusammenhänge zwischen den Gelenkkoordinaten und den kartesischen Koordinaten mit

$$f_i(\mathbf{a}, \mathbf{b}) = {}^{\text{VFP}}\xi_{\text{TCP}} \qquad (3)$$

beschreiben werden. Die Funktionen $\mathbf{f_i}$ können nach [5] aufgestellt werden, welche die Lage des Endeffektors (TCP) zum Mittelpunkt der virtuellen festen Plattform (VFP) beschreiben. Die Berechnung der kartesischen Koordinaten aus gegebenen Gelenkkoordinaten ist für parallele Kinematik durch die Lösung der kinematischen Schleifen möglich [6][7]. Für einen Manipulator mit zwei IKL's können zwei Schleifengleichungen aufgestellt werden, diese sind

$$\mathbf{g}(\mathbf{a}, \mathbf{b}) = \begin{bmatrix} \mathbf{g_1}(\mathbf{a}, \mathbf{b}) \\ \mathbf{g_2}(\mathbf{a}, \mathbf{b}) \end{bmatrix} = \begin{bmatrix} \mathbf{f_2}(\mathbf{a}, \mathbf{b}) - \mathbf{f_1}(\mathbf{a}, \mathbf{b}) \\ \mathbf{f_3}(\mathbf{a}, \mathbf{b}) - \mathbf{f_1}(\mathbf{a}, \mathbf{b}) \end{bmatrix} = \mathbf{0} \qquad (4)$$

Mit $\mathbf{a}$ den primären Variablen beschreiben die Achskoordinaten der Linearachsen $[s_1, s_2, s_3]^T$ und $\mathbf{b}$ als sekundäre Variablen beschreiben die Winkel $[\delta_i, \mu_i]^T$ der Stäbe. Die Endeffektor Pose für den Fall, dass kein Werkzeug montiert ist, liegt deckungsgleich dem Ursprung der virtuellen beweglichen Plattform (VMP). Die virtuellen Plattformen sind durch einen Kreis, welcher durch die Kugelgelenksmittelpunkte verlauft, während die Achsen in ihrer Nullposition stehen, definiert und in Abbildung 2 dargestellt.

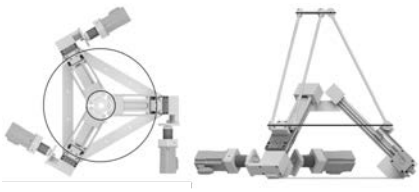


Abbildung 2. Delta 360 Manipulator mit eingezeichneten virtuellen Plattformen.

Die Inverse Kinematik für Parallelkinematik-Roboter ist anders als bei seriellen Manipulatoren eindeutig [5] durch die *Posture Modes of Inverse Kinematic* (PMIK) beschrieben. Mit Hilfe der Gleichungen aus [5]

$$\mathbf{q_k} = \psi({}^{\text{VFP}}\xi_{\text{TCP}})\, für\, i = 1, 2, ..., n_{PMIK(k)} \qquad (5)$$

können die möglichen Lösungen der inversen Kinematik mit der Anzahl n_{PMIK} für die kinematische Kette k ermittelt werden. Nachdem der PMIK bekannt sind, kann die dazugehörige Lösung einfach ausgewählt werden.

Um mögliche Singularitäten steuerungstechnisch vermeiden zu können, muss die Kinematik durch eine Positions- und Geschwindigkeitsanalyse weiter untersucht werden. In dieser kann für den eingesetzten Manipulator bestimmt werden, dass Singularitäten in der vorwärts Kinematik, auf Grund der Antriebsreichweite

nicht erreicht werden. Gleiches gilt für Singularitäten der Inverskinematik.

B. Aufbau von Bewegungssteuerungen für P&P Anwendungen

In dieser Betrachtung von Robotersteuerungen sind Bedieninterface, Prozessablauf und Zustandsdefinition nicht berücksichtigt. Die Funktionen von Interesse beziehen sich auf die unmittelbare Ansteuerung des Roboters, und die dafür notwendigen Bahninterpolationen und Regelung. Um eine P&P Anwendung umsetzten zu können, werden prinzipiell zwei Bewegungsformen benötigt. Einerseits, soll das Annähern an die Werkstücke geradlinig sein, andererseits soll die Bewegungsbahn zwischen den Aufnahme und Ablege Positionen so schnell wie möglich abgefahren werden. Diese Bewegungsanforderungen können durch kartesische Interpolation für die Annäherung, oder durch Punkt zu Punkt (PzP) Interpolation für eine schnellstmögliche Bewegung, umgesetzt werden. Hierfür sind mehrere Parameter zu definieren um eine Trajektorie mit trapezförmigen Geschwindigkeitsprofil zu generieren:

- Startposition $\mathbf{p}_{Start}$
- Zielposition $\mathbf{p}_{Ziel}$
- Geschwindigkeit v
- Beschleunigung a
- Bewegungstyp (kartesisch, PzP)

In Abbildung 3 sind die Positions-, Geschwindigkeits-, und Beschleunigungsprofile für trapezförmige Geschwindigkeitstrajektorien dargestellt. Diese können sowohl für kartesische als auch für Punkt zu Punkt (PzP) Bewegung eingesetzt werden. Je nach dem welcher Raum zur Interpolation genutzt wird, müssen unterschiedliche Randbedingungen zur Berechnung verwendet werden. Eine

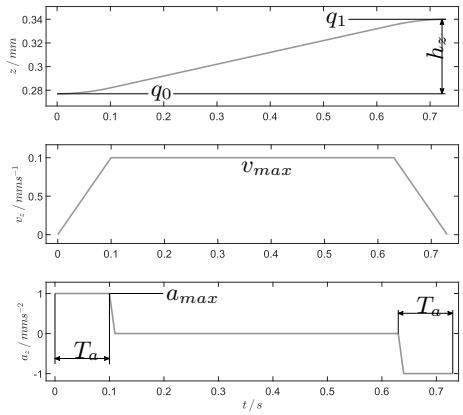


Abbildung 3. Trajektorie mit trapezförmiges Geschwindigkeitsprofil.

genaue Beschreibung zur Herleitung der Interpolationsfunktionen kann in [8] nachgeschlagen werden.

Für (PzP) Bewegungen ergeben sich abhängig der größten Distanz $h_s, i = s_{i,Ziel} - s_{i,Start}$ unterschiedliche Bewegungsdauern T_i. Diese können synchronisiert werden, so dass alle Achsen zur gleichen Zeit zum Stillstand kommen. Dafür werden die Werte der Beschleunigung und Geschwindigkeit auf die größte Bewegungsdauer T_{max} bzw. längste Distanz $h_{s,max}$ nach [8] angepasst durch

$$a_i = \frac{h_i}{T_a(T_{max} - T_a)}, \qquad v_i = \frac{h_i}{T_{max} - T_a}. \tag{6}$$

Wobei die Gelenkkoordinaten s für die Start- und Zielposition durch die inverse Kinematik mit $\mathbf{s} = \psi(\mathbf{p})$ berechnet werden.

Bei der geradlinigen Bewegung ist ähnlich wie zuvor die Distanz $\mathbf{h}_p = \mathbf{p}_{Ziel} - \mathbf{p}_{Start}$

zu ermitteln, jedoch in kartesischen Koordinaten. Durch Normierung dieses Vektors kann die TCP-Geschwindigkeit und Beschleunigung in die drei kartesischen Komponenten zerlegt werden durch

$$\mathbf{v}_{tcp} = ||\mathbf{h}_p|| \cdot v_{tcp}, \qquad \mathbf{a}_{tcp} = ||\mathbf{h}_p|| \cdot a_{tcp}. \quad (7)$$

Durch diese Aufteilung müssen die Basiszeiten nur für eine Komponente bestimmt werden, da alle drei Bahnen die gleichen Zeiten besitzen, solange die Distanz ungleich 0 ist. Die Interpolation erfolgt nun mit den x,y und z Koordinaten durch das trapezförmige Geschwindigkeitsprofil. Zuletzt muss jeder Zeitschritt der Trajektorie durch die inverse Kinematik in Gelenkkoordinaten umgerechnet werden.

C. Anforderungen an die Ausbildungsplattform

Die Anforderungen an die Ausbildungsplattform können grundsätzlich in zwei Teilbereiche aufgeteilt werden. Zum einen sind die Anforderungen zu definieren, welche ermöglichen, dass die Auszubildenden in der Lage sind die Aufgabestellung durchführen zu können. Diese Anforderungen umfassen,

- der Arbeitsablauf muss sinnvoll und verständlich strukturiert werden,
- die Inbetriebnahme der Anlage muss durch geringes Vorwissen von Industriesteuerungen durchführbar sein,
- es dürfen auf keinen Fall Personen verletzt werden,
- der Aufwand, um die Ausbildungsaufgabe durchführen zu können muss angemessen sein,
- die Dokumentation muss vollständig Verfügbar sein,
- eine virtuelle Inbetriebnahme der Steuerung muss möglich sein, und
- die Aufgaben müssen dem theoretischen Wissen angepasst sein.

Die zweite Anforderungsgruppe bezieht sich auf die technische Umsetzung der Ausbildungsaufgabe. Die wichtigsten Anforderungen sind

- die Systemsteuerung muss in einem gängigen Entwicklungsprogramm entwickelt werden,
- die virtuelle und reale Steuerung muss identisch sein,
- die Antriebsachsen dürfen angemessene Geschwindigkeiten nicht überschreiten,
- der Arbeitsbereich der Aufgabe muss durch den Roboter erreichbar sein,
- der Funktionsbereich soll erweiterbar sein,
- die Applikation soll durch ein simples Analysewerkzeug analysiert werden können.

III. Aufbau und Ausbildungsaufgabe

A. Aufbau einer semi-virtuellen P&P Anwendung

Die Ausbildungsaufgabe ist die Entwicklung einer Bewegungssteuerung für eine P&P Anwendung. Um hier auch andere Anwendungsbeispiele wie z.B. sortieren in stationären Boxen oder Entnahme von einem Förderband umsetzten zu können, ohne an der Ausbildungsplattform präsent sein zu müssen, wird ein semi-virtueller Aufbau verwendet, siehe 4. Alle Anwendungsfälle werden mithilfe eines Touchscreens umgesetzt und visualisiert. Somit können die vielseitigen Anwendungen auch für *Remote Laboratories* eingesetzt werden. Die Interaktion zwischen Manipulator und virtueller Anwendung erfolgt durch einen Touchpen. Um den Bildschirm nicht beschädigen zu können,

Abbildung 4. Semi-virtueller Aufbau.

ist der Touchpen federnd gelagert, und eine Sicherheitsfunktion implementiert, welche es nicht erlaubt, dass sich der Roboter durch die Bildschirmebene bewegt.

B. Trajektorien Regelung

Um es zu ermöglichen , dass die Antriebe den vorgegebenen Trajektorien folgen, wird ein Geschwindigkeitsregler mit Positionsrückgabe entworfen. Dieser basiert auf dem *Feedforward* Regelungsprinzip, welches durch seine einfache und effektive Regelung ideal für diese Anwendung ist [9]. Das System wurde mit Hilfe einer Schrittantwort als PT2-System mit

$$\mathbf{A} = \begin{bmatrix} -22,4417 & 0 \\ 1 & 0 \end{bmatrix}, \ \mathbf{B} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \ \mathbf{C} = \begin{bmatrix} 0 & 22,118 \end{bmatrix} \tag{8}$$

identifiziert. Durch Transformation der Übertragungsfunktion in den Zustandsraum und Invertierung der Gleichungen mittels Pseudoinverse kann die Vorsteuerung für den Trajektorieneingang aufgestellt werden. Das Blockdiagramm der Regelung ist in Abbildung 5 gezeigt mit einem PI Regler zum Störgrößenausgleich.

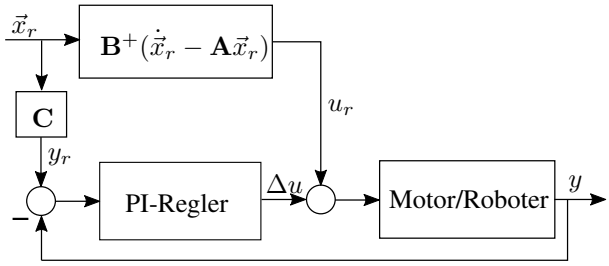


Abbildung 5. Blockdiagramm der *Feedforward* Regelung mit PI-Störgrößenregler($K_p = 10$, $K_i = 0.01$).

C. Ausbildungsaufgabe

Ziel der Aufgabenstellung ist das Entwickeln einer P&P Anwendung durch Verwendung des Delta 360 Manipulators. Die Ausbildungsaufgabe ist so formuliert und strukturiert, dass wichtige Schritte der modellbasierten Entwicklung durchgeführt werden. Der Aufbau der P&P Anwendung besteht aus einem Rüttelbehälter mit zwei unterschiedlichen Bauteilen (Muttern und Zahnräder), zu sehen in Abbildung 6. Die Positionen der Bauteile sind zufällig im Behälter verteilt und werden durch die Pixelkoordinaten an die Steuerung weitergegeben. Ablegepunkte der Bauteile sind zwei Bauteilträger mit jeweils zehn Ablagemöglichkeiten. Aufgabe des Roboters ist es die unsortierten Bauteile aus dem Rüttelbehälter zu entnehmen und sortiert in die Bauteilträger abzulegen.

Damit dieser Prozess gesteuert werden kann wird eine Bahnplanung benötigt mit linearer und Punkt-zu-Punkt (PzP) Bewegung. Die Lineare Bewegung wird bei der Annäherung der Aufnahme und Ablegeposition benötigt. Die Punkt-zu-Punkt Bewegung kann zwischen den Offsetpositionen verwendet werden, womit die Zykluszeit des Roboters gesenkt werden kann. Dafür müssen notwendige Kinematik Funktio-

nen in die Bahnplanung der Systemsteuerung implementiert werden. Zusätzlich benötigt die Bahnplanung echtzeitfähige Interpolationsfunktionen, um die nächste Zielposition der Trajektorie zu berechnen. Um dies realisieren zu können, soll die Validierung der entwickelten Funktionen durch einen virtuellen Manipulator erfolgen. Des Weiteren ist die Montagehöhe des Manipulators zu ermitteln und eine überschlagsmäßige Auslegung der Motorengeschwindigkeiten durchzuführen.

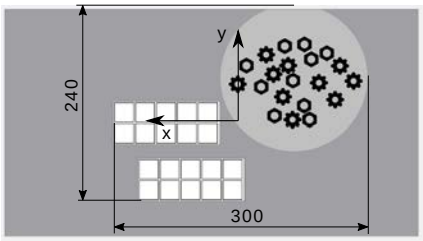


Abbildung 6. Pick und Place Anwendung der Ausbildungsaufgabe.

1. Delta360 CAD-Modellierung

Die Aufgabe beginnt mit der Modellierung eines CAD-Modells des Delta 360 Roboters. Gegeben sind die Bauteile in Unterbaugruppen zusammengefügt. Diese müssen durch Abhängigkeiten verbunden werden. Die Auswahl der Abhängigkeiten ist so zu treffen, dass das Modell die korrekten kinematischen Eigenschaften besitzt.

2. Delta360 *Simscape*-Modellierung

Das generierte CAD-Modell muss in *Simscape* importiert und aufbereitet werden. Dafür steht das *Simscape Multibody Plugin* für *Autodesk Inventor* zur Verfügung. Als Austauschformat wird eine XML-Datei aus der Baugruppe mit den dazugehörigen STEP-Dateien generiert. Diese soll in MATLAB eingelesen und in ein Mehrkörpermodell umgewandelt werden. Für die weitere Verwendung müssen Achsrichtungen, Endeffektorpositionsmessung und Solvereinstellungen angepasst werden. Des Weiteren müssen Bewegungsbefehlblöcke erzeugt und integriert werden, so dass die Linearachsen angetrieben werden können.

3. Applikationsanalyse

Um passende Geschwindigkeiten und Arbeitsbereiche definieren zu können, soll die Applikation analysiert werden. Dafür kann die zur Verfügung gestellten Toolbox verwendet werden. Modifikationen bzw. Erweiterungen der Toolbox können entwickelt werden um ein genaueres Analyseergebnis zu bekommen bzw. um die Verwendung der Toolbox zu vereinfachen. Ziel der Analyse ist die Bestimmung der notwendigen Achsgeschwindigkeiten, um die vorgegebenen zehn Zyklen pro Minute zu erreichen, und die Ermittlung der Montagehöhe um den gegebenen Arbeitsbereich der Anwendung abzudecken.

4. Generieren Vorwärts- und Inverskinematik Funktionen

Erster Schritt für die Modifizierung des Pfadplanungsblockes ist die Entwicklung eigener Vorwärts- und Inverskinematik Funktionen sowie die Herleitung der Jocobimatrix der Inverskinematik als Grundlage der Geschwindigkeitstransformationsfunktion. Diese können mit Hilfe der Berechnungen im Anhang II-A generiert werden. Nachher sollen sie in die Steuerung eingebunden und getestet werden.

5. Generieren Interpolationsfunktionen für lineare und PzP Bewegungen

Der zweite Schritt der Modifikation ist, dass zwei Funktionen zur Interpolation der Bewegungen aufgestellt werden. Die erste Funktion dient der Achsrauminterpolation und die zweite

der Arbeitsrauminterpolation. Wie bereits angesprochen werden für eine *P&P* Anwendung beide Bewegungstypen benötigt.

6. Inbetriebnahme des virtuellen Roboters mit eigener Systemsteuerung

In diesem Arbeitsschritt sollen alle generierten Funktionen anhand der Systemsteuerung mit eigens entwickelten Mehrkörpermodell getestet werden. Somit können Fehler ausgebessert und Optimierungen durchgeführt werden, ohne dass es zu Schäden der Anlage kommen kann.

7. Inbetriebnahme der realen Steuerung

Sobald alle Funktionen auf der virtuellen Steuerung getestet wurden, können diese in die reale Anlage integriert werden. Die Inbetriebnahme umfasst das Überspielen der entwickelten Funktionen auf die *Simulink*-Steuerung am IPC der Anlage. Anschließend muss die Steuerung der SPS aktiviert werden und die Schnittstelle zwischen *Simulink* und TwinCAT verbunden werden. Sofern diese Tätigkeiten durchgeführt wurden, müssen die Anwendungsparameter in der HMI angepasst, und anschließend kann die *P&P* Anwendung getestet werden.

IV. Software & Hardwareimplementierung

Zur Implementierung von Kinematik- und Interpolationsfunktionen wird eine Robotersteuerung mit ähnlicher Architektur wie in [10] entwickelt. Die Steuerungsarchitektur ist in Abbildung 7 zu sehen. Eingaben vom Bediener erfolgen über die HMI Schnittstelle, wodurch zum Beispiel die Applikation gestartet oder gestoppt werden kann. Die zentrale Logik der Steuerung ist die Applikationssteuerung, in welcher der Zustand des Roboters definiert und die P&P Anwendung abgearbeitet wird. Ausgang dieser Steuerung sind die Bewegungsbefehle an die Bewegungssteuerung. Diese Befehle enthalten Daten wie zum Beispiel Zielpositionen und Bewegungseigenschaften, wodurch die Bewegungssteuerung den Roboter gezielt antreibt. Damit Endeffektor oder Peripheriegeräte gesteuert werden können, ist eine E/A Steuerung integriert. Die Überwachung des sicheren Zustandes erfolgt durch die Sicherheitssteuerung, welche bei Aktivierung der Sicherheitsfunktionen die Motoren stromfrei schaltet und die Applikationssteuerung in einen Fehlermodus zwingt. Diese eigenentwickelte Steuerung ermöglicht es die Funktionen der Bewegungssteuerung aus II durch die entwickelten Funktionen der Auszubildenden zu ersetzten. Sicherheitsfunktionen bleiben weiterhin funktionsfähig, wodurch eine sichere Implementierung und Inbetriebnahme sichergestellt ist.

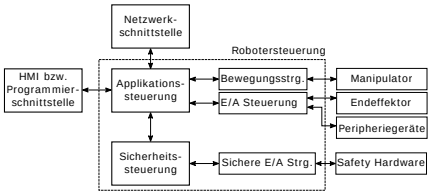


Abbildung 7. Architektur der Robotersteuerungskomponenten.

A. Integration der Funktionen

Die Berechnungen und der dahinterstehende Code der Kinematik- und Interpolationsfunktionen sind aus der Toolbox bekannt. Einzelne Codesegmente können deshalb eins zu eins übernommen werden. Alle Segmente mit Schleifen bzw. Vektoren für den zeitlichen Verlauf mit n Einträgen müssen in abgeänderter Form übernommen werden. D.h. Schleifen über die Zeit werden aus der Berechnung entfernt, und Vektoren mit n Einträgen werden auf 1 reduziert.

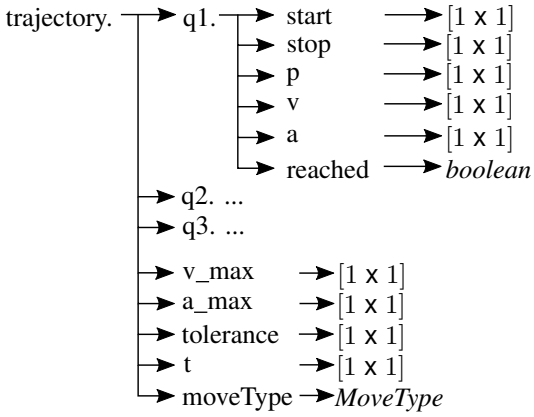


Abbildung 8. Diagramm der Trajektoriestruktur.

Um die Daten des vorhergehenden Interpolationsschrittes zu speichern, werden persistente Variablen verwendet. Für die Funktionen der Vorwärts- bzw. Inverskinematik müssen keinerlei Änderungen durchgeführt werden, um sie in der Steuerung einsetzen zu können.

Zu Beginn der Integration müssen die Funktionseingänge und -ausgänge der verschiedenen Funktionen geklärt werden. Nachdem die Datensignale zwischen den *Simulink*-Blöcken durch Busstrukturen aufgebaut sind, müssen die Strukturen der relevanten Signale bekannt sein. Die Trajektorienstruktur ist in Abbildung 8 dargestellt. Diese wird innerhalb der *MoveJ* und *MoveL* Bereiche verwendet und ist somit für die Inverskinematik, als auch für die Interpolationsfunktionen relevant. Welchen Raum die Daten widerspiegeln, hängt also von der vor- bzw. nachgeschalteten Inverskinematik ab. Wann die Inversekinematik zum Einsatzt kommt, ist durch den Bewegungstyp in der Variable *MoveType* definiert.

Die zweite relevante Struktur ist die *Jointspace* bzw. *Taskspace* Struktur. Hier werden die vom Roboter zurückgeführten Achspositionen durch die Vorwärtskinematik in Endeffektorpositionen umgerechnet. Dies ist notwendig, um zu überwachen, ob der Endeffektor seine Zielpostion erreicht hat, um somit den nächsten Wegbefehl in der Bewegungssteuerung abzuarbeiten. Die Strukturen besitzen den gleichen Aufbau wie in Abbildung 9 dargestellt, nur die Werte beschreiben unterschiedliche Koordinaten.

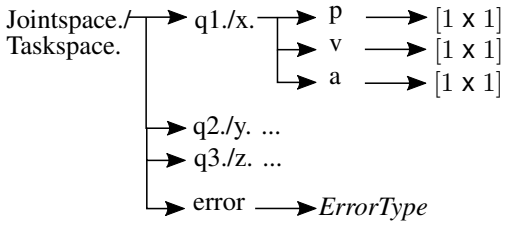


Abbildung 9. Diagramm der *Jointspace* bzw. *Taskspace* Struktur.

B. Lineare Interpolation

Die hier verwendete Trajektorienstruktur beinhaltet alle relevanten Daten, um die Interpolation durchzuführen. Jedoch werden in den einzelnen Funktionen nicht alle Daten benötigt, weshalb je nach Funktionstyp unterschiedliche Daten aus der Struktur zur Berechnung verwendet, und die Ergebnisse der Funktion in die Struktur geschrieben werden. Somit können die benötigten Daten und berechneten Ergebnisse der Interpolation im kartesischen Raum aus Tabelle II gelesen werden.

Tabelle II
EINGÄNGE UND ERGEBNISSE DER LINEAREN TRAJEKTORIEBERECHNUNG.

Funktionseingang	Beschreibung	Einheit
$p_{start(x,y,z)}$	kartesischer Startpunkt	m
$p_{stop(x,y,z)}$	kartesischer Zielpunkt	m
v_{max}	maximale Geschwindigkeit	$\mathrm{m\,s^{-1}}$
a_{max}	maximale Beschleunigung	$\mathrm{m\,s^{-2}}$
t	Zeit seit Trajektorienstart	s
MT	Bewegungstyp	*MoveType*
Ergebnisse	Beschreibung	Einheit
$p_{(x,y,z)}(t)$	interpolierte Position	m
$v_{(x,y,z)}(t)$	interpolierte Geschw.	$\mathrm{m\,s^{-1}}$
$a_{(x,y,z)}(t)$	interpolierte Beschl.	$\mathrm{m\,s^{-2}}$

Tabelle III
EINGÄNGE UND ERGEBNISSE DER PZP TRAJEKTORIEBERECHNUNG.

Berechnungseingang	Beschreibung	Einheit
$s_{start(1,2,3)}$	Startpunkt Achse	m
$p_{stop(x,y,z)}$	kartesischer Zielpunk	m
v_{max}	maximale Geschwindigkeit	$\mathrm{m\,s^{-1}}$
a_{max}	maximale Beschleunigung	$\mathrm{m\,s^{-2}}$
t	Zeit seit Trajektorienstart	seconds
MT	Bewegungstyp	MoveType
Ergebnisse	Beschreibung	Einheit
$s_{(1,2,3)}(t)$	interpolierte Position	m
$\dot{s}_{(1,2,3)}(t)$	interpolierte Geschw.	$\mathrm{m\,s^{-1}}$
$\ddot{s}_{(1,2,3)}(t)$	interpolierte Beschl.	$\mathrm{m\,s^{-1}}$

Nach dieser Berechnung erfolgt die Transformation in den Achsraum, um den Roboter steuern zu können. Dafür werden die kartesischen Position $p_{(x,y,z)}(t)$ durch die Inverskinematik in Achskoordinaten $s_{(1,2,3)}(t)$ umgewandelt. Ebenfalls können die kartesischen Geschwindigkeiten $v_{(x,y,z)}(t)$ mittels der Jacobimatrix der Inverskinematik in Achsgeschwindigkeiten $\dot{s}_{(1,2,3)}(t)$ umgerechnet werden. Durch Befüllen der *Jointspace* Struktur können die Daten an die Motorsteuerung des Roboters weitergegeben werden, um den nächsten Punkt auf der linearen Trajektorie anzufahren.

C. Achsraum Interpolation

In der Achsraum Interpolation erfolgt eine umgekehrte Reihenfolge der oben verwendeten Funktionen. Zu Beginn werden die Start- und Zielpunkte vom Arbeitsraum in den Achsraum transformiert. Die maximalen Geschwindigkeits- und Beschleunigungsangaben sind in diesem Fall maximale Achsgeschwindigkeiten bzw. Achsbeschleunigungen. Anschließend werden die nächsten Achsraumwerte durch die Interpolationsfunktion berechnet. Tabelle III listet die relevanten Funktionseingänge und Ergebnisse auf.

V. DURCHFÜHREN DER AUSBILDUNGSAUFGABE

Zu Beginn der Aufgabendurchführung muss definiert werden, welche Daten, Dokumentationen und Softwarezustände den Auszubildenden zur Verfügung gestellt werden. Die für die Umsetzung des CAD Modells notwendigen Datenblätter und Stücklisten des Manipulators können von Igus direkt bezogen werden. Bedienanleitungen für *Simscape* und den *Simscape Multibody Link* können bei Mathworks bezogen werden. Des Weiteren erhalten die Auszubildenden die komplette Toolbox mit Quellcode um primäre Analysen durchführen zu können. Sollten während der Arbeiten Optimierungsmöglichkeiten identifiziert werden, so können diese mit Hilfe der Quellcodedateien umgesetzt und getestet werden. Für die Umsetzung der virtuellen und realen Inbetriebnahme wird die Systemsteuerung bereitgestellt, jedoch ohne *Simscape*-Modell. Zuletzt müssen die Strukturen und Variablen innerhalb der Bewegungssteuerung bekannt sein, so dass diese richtig durch ihre eigenentwickelten Funktionen bearbeitet werden können. Beispiele für diese Dokumentation sind die Abbildungen 8 und 9.

Die Aufgabe 1 ist mit Hilfe des *Simscape Multibody Link Manuel & User Guide* ohne größere Schwierigkeiten durchzuführen. Es müssen nur zwei Punkte näher erläutert werden, ohne welche viel Aufwand betrieben werden muss, um die resultierenden Fehler zu lösen. Zum einen ist in Inventor die Verwendung der Funktion *Verbindungen* zu vermeiden, da die Abhängigkeiten und Freiheitsgrade mit dieser Funktion bei komplexeren Strukturen, wie dem Delta 360, nicht mehr korrekt übersetzt werden können. Andererseits ist darauf aufmerksam zu machen, dass eine deutsche Inventor Version ein Komma als Dezimalseparator verwendet. Damit kann MATLAB nicht umgehen und der Importprozess wirft Fehler auf. Dieser Fehler ist schlecht beschrieben und es ist nicht verständlich wie dieser behoben werden kann.

In der Aufgabe 2 werden die Auszubildenden zum ersten Mal das Identifikationsproblem der Kinematikketten sehen. Hier muss eine kurze Beschreibung zur Verfügung stehen, in der die Schritte der Identifikation der Kinematikketten beschrieben wird. Ansonsten können *Move Demand* Blöcke integriert, und eine Transformationsmessung zur Bestimmung der Endeffektorposition erstellt werden. Nach diesen Schritten ist das Simulationsmodell einsatzbereit.

Für Aufgabe 3 müssen überschlagsmäßige Berechnungen der notwendigen Achsgeschwindigkeiten, und der optimalen Montagehöhe des Manipulators durchgeführt werden. Dafür kann die zur Verfügung gestellte Toolbox verwendet werden. Zur Ermittlung der Geschwindigkeiten können die durchschnittlichen Distanzen der P&P Aufgabe in der *Task*-Klasse simuliert werden. Die entstehenden Simulationszeiten werden anschließend in Zykluszeiten umgerechnet und mit den vorgegebenen zehn Zyklen pro Minute verglichen. Durch Wiederholen dieser Analyse mit unterschiedlichen Geschwindigkeiten, können die Geschwindigkeiten gezielt ausgewählt werden. Da die Beschleunigungswerte ebenfalls Einfluss auf dieses Ergebnis nehmen, können diese als Faustregel mit zwei bis dreimal die Geschwindigkeitswerte angenommen werden. Jedoch sollten Beschleunigungen größer als $0.2\,\mathrm{m\,s^{-2}}$ vermieden werden, da der Regler und die Motoren ansonsten überlastet werden. Um die vorgegebenen zehn Zyklen pro Minute abfahren zu können, ergibt sich eine Mindestgeschwindigkeit von $0.09\,\mathrm{m\,s^{-1}}$ bei einer Beschleunigung von $0.18\,\mathrm{m\,s^{-2}}$ heraus.

Um die Aufgabe 4 umsetzten zu können, müssen eigenentwickelten Funktionen der Vorwärts- und Inverskinematik, sowie die Geschwindigkeitstransformationsfunktion zwischen kartesischen Raum und Achsraum in die vorhergesehenen externen Funktionen integriert werden. Dabei ist auf die Datenstrukturen aus Kapitel IV-A zu achten. Diese sind bereits in den Platzhaltefunktionen zu sehen. Prinzipiell müssen in der Inversekinematikfunktion die Daten der Trajektorienstruktur des Stopppunkts und des aktuellen Punkts, sowie die aktuellen Geschwindigkeiten transformiert werden. Bei der Vorwärtskinematik können die zurückgeführten Daten des Manipulators einfach in den kartesischen Raum transformiert werden.

In Aufgabe 5 werden die Bewegungsfunktionen in die Steuerung integriert. Dabei kann für beide Bewegungsformen die gleiche Interpolationsfunktion für trapezförmige Geschwindigkeitsprofile verwendet werden. Nur die Berechnung der Randbedingungen muss durch unterschiedliche Berechnungen erfolgen, je nach

Bewegungsart. Wie diese Berechnungen aufgestellt sind kann direkt aus dem Quellcode entnommen werden.

Die Aufgabe 6 ist das Simulieren der eigenentwickelten Funktionen, um diese zu validieren. Für einen effizienteren Validierungsprozess, können die externen Funktionen einzeln getestet werden. Dadurch ist die Fehlersuche um vieles einfacher, als beim Testen der wesentlich komplexeren Gesamtsteuerung.

In der Aufgabe 7 wird die reale Anlage in Betrieb genommen. Dazu müssen die externen Funktionen mit Hilfe eines Datenträgers auf die Hardware der Ausbildungsplattform gespeichert werden. Auf der Ausbildungsplattform ist die identische Steuerung gespeichert und kann daher gleich wie in der virtuellen Steuerung betrieben werden. Der einzige Unterschied ist, dass die SPS-Steuerung bei einem Neustart der Anlage kompiliert werden muss. Dazu ist *TwinCAT* zu öffnen und das Programm *Ausbildungsplattform_Agent.sln* muss geladen und durch *Konfiguration aktivieren* kompiliert werden. Im Anschluss kann das MATLAB-File *RunAusbildungsplattform.m* ausgeführt werden und mit der *TE 1410* Variante gestartet werden. Nun ist der Roboter einsatzbereit und kann über die HMI betrieben werden.

VI. Zusammenfassung & Ausblick

Abschließend werden die einzelnen Punkte der Aufgabenstellung mit den Resultaten der Arbeit verglichen, um sicherzustellen, dass alle relevanten Anforderungen zur Durchführung einer Robitiklehre anhand der Ausbildungsplattform umgesetzt wurden. Im zweiten Teil dieses Kapitels wird Aufgezeigt, welche weiteren Potentiale zur Ausbildung in der Robotik noch vorhanden sind, und welche erweiterten Themengebiete mit der Plattform erschlossen werden können.

Durch die Anforderungsanalyse wird aufgezeigt, welche Funktionalitäten für die Ausbildungsplattform gegeben sein müssen, um eine praxisnahe Ausbildung für Analyse, Entwicklung und Umsetzung von Robotersystemen durchzuführen. Aus diesen Erkenntnissen kann ein ganzheitliches Konzept der Plattform mit virtuellen und realen Komponenten, sowie passenden Schnittstellen aufgestellt werden. Die Umsetzung dieses Konzepts kann in vier Hauptpunkte aufgeteilt werden. Beginnend mit der Ausbildungsaufgabenformulierung, gefolgt von der Hardwareauswahl und dem Hardwaredesign, der Toolboxentwicklung mit Simulationsmodell, und der abschließenden Steuerungsentwicklung.

Die Ausbildungsaufgabe ist modular aufgebaut, so dass die einzelnen Aufgaben unabhängig voneinander durchgeführt werden können. Durch die modulare Aufgabenstellung ist es immer möglich den realen Roboter inbetriebzunehmen, und somit ein wirkliches Erfolgserlebnis zu generieren. Als Ausgangspunkt der Aufgabe dienen die HMI und MATLAB Dateien mit blanken externen Kinematik und Bewegungsfunktionen, sowie dem Vorlage *Simulink*-Bibliotheksblock des Simulationsmodells. Ebenfalls werden die CAD-Daten des Manipulators in Form von Baugruppen der beweglichen Komponenten zur Verfügung gestellt.

Der semi-virtuelle Aufbau ist in Abbildung 4 dargestellt und kann die Anforderungen an den Bauraum, und die Abgrenzung des Manipulators nach außen gewährleisten. Die getroffene Auswahl der Hardwarekomponenten stellt sicher, dass beide Displays gleichzeitig durch den IPC betrieben, und die notwendigen Software-

systeme gleichzeitig verwendet werden können. Eine angemessene Sicherheit ist durch die Sicherheitskomponenten von Pilz und Beckhoff gegeben.

Um es den Auszubildenden zu ermöglichen, eigenentwickelte Funktionen der Steuerung in die reale Anlage implementieren zu können, ist eine *Simulink* basierte Robotersteuerung entworfen worden. Der Quellcode und die Bibliotheken der gesamten Steuerung werden den Auszubildenden zur Verfügung gestellt, wodurch es ihnen ermöglicht wird, ihre Funktionen zu implementieren und virtuell am eigenen Simulationsmodell zu testen. Nach erfolgreichem verwenden des virtuellen Modells kann die Steuerung auf den IPC der Anlage übertragen werden. Die reale Steuerung ist so aufgesetzt, dass eine Kompilierung in *Simulink* genügt um den Roboter zu steuern. Dafür ist eine Kommunikationsschnittstelle zwischen MATLAB und TwinCAT integriert. Die TwinCAT-Steuerung bekommt die Bewegungsdaten aus *Simulink* und steuert den Roboter über einen integrierten PI-Regler. Eine weitere Variante zur Übertragung der, von den Auszubildenden entwickelten, Steuerung auf das reale System ist mittels automatischer Codegenerierung, wie in [3] dokumentiert. Diese Methode ist jedoch komplizierter und erfordert vertieftes wissen in der Automatisierungstechnik mit TwinCAT-Steuerungen. Somit kann diese Variante, obwohl leistungsfähiger als die *Simulink-TwinCAT* Schnittstelle, nicht als Standardaufgabe gestellt werden, da der Fokus von Robotik auf Automatisierungstechnik verschoben wird.

Ein weiteres Zukunftsziel der Anlage ist die Einsatzmöglichkeit in Remote-Lab Applications. Dafür müssen Kameras in der Anlage befestigt und Netzwerkschnittstellen definiert werden. Die Verbindung auf den IPC und die Videodaten sollte über eine Benutzeroberfläche funktionieren, sodass die Arbeit während der Online-Inbetriebnahme übersichtlich bleibt und die Auszubildenden sich auf das Wesentliche konzentrieren können.

LITERATUR

[1] T. Brogardh, *PKM Research - important Issues, as seen from a Produkt Development Perspective at ABB Robotics*. Sweden: ABB Automation Technology / Robotics, 2002. [Online]. Available: https://robot.gmc.ulaval.ca/fileadmin/documents/images/workshop2002/BROGARDH_ABSTRACT.pdf

[2] N. S. Tlale and P. Zhang, "Teaching the design of parallel manipulators and their controllers implementing matlab, simulink,simmechanics and cad," *Int. J. Engng*, vol. 21, no. 5, pp. 838–845, 2005.

[3] J. Muigg, "Entwicklung einer Ausbildungsplattform zum modellbasierten Entwicklungsprozess von Automatisierungslösungen am Beispiel von Industrierobotenkinematiken." Master Thesis, MCI — Die unternehmerische Hochschule, 2021.

[4] Manfred Husty, Adolf Karger und Hans Sachs, *Kinematik und Robotik*.

[5] M. K. Özgören, *Kinematics of general spatial mechanical systems*. Hoboken, NJ: John Wiley & Sons Inc, 2020.

[6] H. G. J. He and Z. Wang, *Solving the forward kinematics problem of six-DOF Stewart platform using multi-task Gaussian process.* Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science, 227(1), 161-169, 2013.

[7] G. Liu, "Optimal kinematic design of a 6-ucu kind gough-stewart platform with a guaranteed given accuracy." *Robotics*, vol. 7, no. 2, p. 30, 2018.

[8] L. Biagiotti and C. Melchiorri, *Trajectory Planning for Automatic Machines and Robots*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009. [Online]. Available: http://dx.doi.org/10.1007/978-3-540-85629-0

[9] P. Lambrechts, M. Boerlage, and M. Steinbuch, "Trajectory planning and feedforward design for electromechanical motion systems," *Control Engineering Practice*, vol. 13, no. 2, pp. 145–157, 2005.

[10] A. Pott and T. Dietz, *Industrielle Robotersysteme: Entscheiderwissen für die Planung und Umsetzung wirtschaftlicher Roboterlösungen*, 2019.

Serafin Kollegger ist Laboringenieur im Studiengang Mechatronik am MCI Innsbruck/Österreich. Er Unterrichtet in den Kompetenzbereichen der Konstruktion und Produktentwicklung, und unterstützt Forschungsprojekte in der Realisierungsphase.

DQ-Sensor-Less Control Implementation on a Multi-Phase High-Speed Direct Drive Motor

Martin Maier, and Maurizio Incurvati (supervisor)

***Abstract*—Modern motor drives are not only power electronics. Advanced control techniques need to be implemented into microcontrollers in order to achieve high-level performances. In particular, detection of position and speed is a topic of primary importance. In this work, a sensor-less control technique for a high-speed multi-phase axial flux motor is developed. It is implemented and tested in an automotive grade microcontroller. For this purpose, a field-oriented control is designed. This is done with a cascaded control structure with an inner current loop and an outer speed loop. For the controller of this, PI controllers are used. The back-EMF of the motor is estimated using a model-based control method. For this purpose, a full-order sliding mode flux observer (SMO) is designed. For the rotor speed and angle estimation an improved quadrature phase-locked loop (PLL) is used. The system with the designed controllers is implemented in Matlab/Simulink. For a realistic simulation all relevant parameters of the circuit, ADCs, data acquisition and calculation are considered. Furthermore, the controllers are implemented on an Aurix microcontroller and the simulations are evaluated and compared with measurements on the real system. When evaluating the measurement results, the system shows a very robust behavior over a wide speed range.**

***Index Terms*—High-speed multi-phase axial flux motor, cascaded PI-controller, full order sliding mode flux observer, improved quadrature phase-locked loop, Aurix microcontroller**

I. Introduction

AXIAL flux permanent magnet synchronous motors (AFPMSMs) have the advantage over radial permanent magnet synchronous motors (PMSMs) in that they have a compact axial structure and high power density [1]. In the controller design, the techniques of PMSM can be used.

M. Maier study at the MCI, e-mail: martin-maier94@gmx.at.

The system is usually a combination of an inverter and the motor itself, what has high efficiency and, with a suitable controller, high performances in terms of position, speed or torque control. A commonly used control technique is the field-oriented control (FOC), which is essentially a cascaded controller in the rotor flux reference plane, also called dq-plane (direct quadrature) [2].

To control the motor in the dq-plane, the controller needs accurate feedback of the rotor position and speed for the speed control. The sensors for this can be expensive and bulky, and especially at high speeds, the manufacturing of the motor can lead to inaccuracies in the position and speed measurements [3], [4].

Since mechanical sensors are unwanted, sensorless control techniques are widely used. Sensorless control techniques can generally be divided into low-speed methods and high-speed methods. Low-speed methods use high frequency injection to calculate motor position via motor reluctance. Model-based methods are used to calculate position and velocity at high-speeds. This often involves estimating the back-EMF or flux of the motor, from which the speed and angle can be calculated. This can be done with a sliding mode observer (SMO), an extended Kalman filter (EKF), or a Luenberger state observer, among others. [3], [5].

This paper presents the design of cascaded PI controllers for the FOC using common tools such as pole-zero cancellation and loop shaping technique. In addition, a design method for a full-order sliding

mode controller is proposed, where the switching gain is designed with the Lyapunov stability theorem and observer poles are selected when the sliding mode occurs. A disadvantage of this method is the chatter phenomenon that occurs, which comes from the switching function that pushes the estimated current onto the sliding surface. The undesirable chattering phenomenon is reduced using a saturation function instead of the signum function. Using the estimated back-EMF of the observer, the speed and position of the motor are calculated with a quadrature phase-locked loop. The system is simulated with the ADC delays and the SPVM update delays. To evaluate the system, the simulation is compared with measurements of the real system, where the motor control is implemented on an Aurix microcontroller.

II. System overview

A block diagram of the overall system with the FOC, the speed and angle estimation, and the motor and its inverter is shown in Fig. 1.

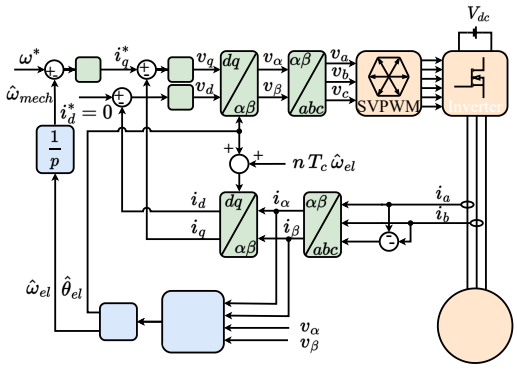


Fig. 1. Block diagram of the control algorithm including the FOC and the sensorless SMO based control technique.

III. Field-Oriented Control

The field-oriented control (FOC) approach is used for the speed control of the system. An overall block diagram of this approach is shown in Fig. 2. From the block diagram, it can be seen that there are two PI controllers nested in two loops. The outer loop is the speed loop and the inner loop is the current (torque) loop.

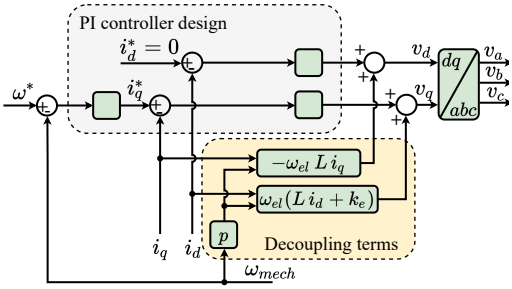


Fig. 2. Block diagram of the FOC with the decoupling network.

For the controller designs, the differential equations of the currents must be transformed into the dq-coordinate system. For this purpose, the AFPMSM can be considered as a non salient permanent magnet synchronous motor (PMSM) with a sinusoidal back EMF. The differential equation for the currents is thus

$$\begin{aligned}\dot{i}_a(t) &= -\frac{R}{L}\,i_a(t) - \frac{1}{L}\,e_a(t) + \frac{1}{L}\,v_a(t)\\ \dot{i}_b(t) &= -\frac{R}{L}\,i_b(t) - \frac{1}{L}\,e_b(t) + \frac{1}{L}\,v_b(t)\\ \dot{i}_c(t) &= -\frac{R}{L}\,i_c(t) - \frac{1}{L}\,e_c(t) + \frac{1}{L}\,v_c(t),\end{aligned} \tag{1}$$

where $e(t)$ is the back electromotive force induced by the rotation of the permanent magnet rotor, R is the stator resistance, L is the inductance of the motor and $v(t)$ are the phase voltages [2], [6].

Applying the Clarke-Park transformation [2], [5], [7] to Eqn. 1 yields

$$\begin{aligned}\dot{i}_d &= \frac{1}{L}\,v_d - \frac{R}{L}\,i_d + \omega_{el}\,i_q\\ \dot{i}_q &= \frac{1}{L}\,v_q - \frac{R}{L}\,i_q - \omega_{el}(i_d + \frac{K_e}{L})\end{aligned} \tag{2}$$

for the currents in the dq-plane. In Eqn. 2, additional cross-coupling terms appear due to the Clarke-Park

transformation. These terms include the electric angular speed ω_{el} and the back-EMK constant K_e. [2], [6].

For PI controller design, this cross-coupling term can be neglected, but at a later stage of the design it can be added as a feed-forward term as shown in Fig. 2. [6], [7].

A cascaded control structure is used for speed control. Two control loops are used for this purpose, an outer speed control loop and an inner current control loop. Both control loops are controlled by PI controllers. An advantage of the FOC is that the same PI controller can be used for flux and torque control (dq currents) [6], [8].

The controller of the inner current loop is designed by the pole-zero cancellation method and the controller of the outer speed loop is designed with the loop-shaping technique. These methods are commonly used in automatic control and are given in [8], [9]. Since this is a cascaded controller, the bandwidth of the outer loop must be at least 10 times smaller than the bandwidth of the inner loop.

IV. Full-Order Sliding Mode Observer

The current equation of all three phases is given in Eqn. 1. Transforming this equation with the Clarke transformation gives the state equation in the $\alpha\beta$-coordinate system.

$$\dot{\vec{i}}_{\alpha\beta}(t) = \boldsymbol{A}\vec{i}_{\alpha\beta}(t) + \boldsymbol{B}(\vec{v}_{\alpha\beta}(t) - \vec{e}_{\alpha\beta}). \quad (3)$$

The back-EMF based on the active flux linkage is

$$\vec{e}_{\alpha,\beta}(t) = K_e\,\omega_{el}\begin{bmatrix}-\sin(\theta_{el})\\ \cos(\theta_{el})\end{bmatrix}. \quad (4)$$

Thus, the circuit equation in $\alpha\beta$-coordinates has a nonlinear term containing trigonometric functions. Assuming that the mechanical time constant of the system is much larger than the electromagnetic one, the angular speed can be considered as a constant in one control cycle $\dot{\omega}_{el} = 0$. Moreover, the property that the back-EMF constant K_e is not time dependent can be used to linearize the circuit equation [4], [5], [10].

By extending the circuit equation with the back-EMF yields the full order state equation

$$\begin{aligned}\dot{\vec{x}} &= \begin{bmatrix}\boldsymbol{A_{11}} & \boldsymbol{A_{12}}\\ \boldsymbol{0} & \boldsymbol{A_{22}}\end{bmatrix}\vec{x} + \begin{bmatrix}\boldsymbol{B_1}\\ \boldsymbol{0}\end{bmatrix}\vec{v}_{\alpha\beta}\\ \vec{y} &= \boldsymbol{C}\,\vec{x},\end{aligned} \quad (5)$$

where

$$\begin{aligned}\vec{x} = [i_\alpha\ i_\beta\ e_\alpha\ e_\beta]^{\mathrm{T}},\ \vec{y} = [i_\alpha\ i_\beta]^{\mathrm{T}},\\ \boldsymbol{A_{11}} = -(R/L)\mathbf{I},\ \boldsymbol{A_{12}} = -(1/L)\mathbf{I},\\ \boldsymbol{A_{22}} = \omega_{el}\mathbf{J},\ \boldsymbol{B_1} = (1/L)\mathbf{I},\\ \mathbf{I} = \begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix},\ \mathbf{J} = \begin{bmatrix}0 & -1\\ 1 & 0\end{bmatrix}.\end{aligned} \quad (6)$$

The sliding mode observer is based on the principle of a disturbance observer, however due to the extension of state space model with the back-EMF it can be categorized as full-order observer. For the sliding mode observer first the sliding surface has to be defined. Therefore, the state variables of the current are used. Thus, the sliding hyperplane is $\vec{S} = \hat{\vec{i}}_{\alpha\beta} - \vec{i}_{\alpha\beta} = \tilde{\vec{i}}_{\alpha\beta}$. With the state Eqn. 5 the full-order sliding mode control law can be established as

$$\begin{aligned}\dot{\hat{\vec{i}}} = \begin{bmatrix}\boldsymbol{A_{11}} & \boldsymbol{A_{12}}\\ \boldsymbol{0} & \hat{\boldsymbol{A}}_{\boldsymbol{22}}\end{bmatrix}\hat{\vec{i}} + \begin{bmatrix}\boldsymbol{B_1}\\ \boldsymbol{0}\end{bmatrix}\vec{v}_{\alpha\beta}\\ - \frac{1}{L}\,\boldsymbol{K}\,\mathrm{sgn}(\vec{S}),\end{aligned} \quad (7)$$

where $\hat{}$ donates the estimated values and

$$\begin{aligned}\hat{\boldsymbol{A}}_{\boldsymbol{22}} &= \hat{\omega}_{el}\,\mathbf{J},\\ \boldsymbol{K} &= \begin{bmatrix}k_S & 0\\ 0 & k_S\\ -m & 0\\ 0 & -m\end{bmatrix},\end{aligned} \quad (8)$$

with k_S as switching gain and m as observer poles [4], [5], [11].

Fig. 3 shows the block diagram of the observer.

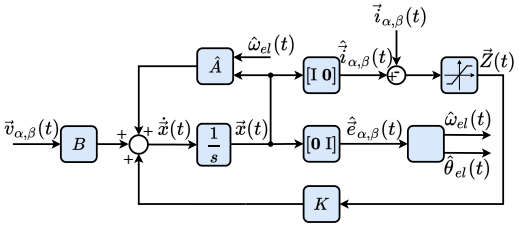


Fig. 3. Block diagram of the full order sliding mode observer.

A. Switching Gain

Determining the switching gain k_S is done with the help of the Lyapunov stability analysis [4]. Therefor the Lyuaponov function is

$$V = \frac{1}{2}\,\vec{S}^{\mathrm{T}}\,\vec{S} = \frac{1}{2}\,(\tilde{i}_\alpha^2 + \tilde{i}_\beta^2). \tag{9}$$

For sure $V > 0$, then with the stability decision theorem the observer reaches a stable state if $\dot{V} < 0$. The time derivative of the Lyapunov function is

$$\dot{V} = \dot{\vec{S}}^{\mathrm{T}}\,\dot{\vec{S}} = \dot{\tilde{i}}_\alpha\,\dot{\tilde{i}}_\alpha + \dot{\tilde{i}}_\beta\,\dot{\tilde{i}}_\beta. \tag{10}$$

With the error equation build by subtracting Eqn. 5 from Eqn. 7 and substituting this in Eqn. 10 leads to

$$\begin{aligned}\dot{V} = &-\frac{R}{L}(\tilde{i}_\alpha^2 + \tilde{i}_\beta^2) \\ &+ \frac{1}{L}\left[\tilde{i}_\alpha(\tilde{e}_\alpha + k_S z_\alpha) + \tilde{i}_\beta(\tilde{e}_\beta + k_S z_\beta)\right],\end{aligned} \tag{11}$$

where $z = \mathrm{sgn}(\tilde{i})$ is the switching function. From this it can be seen that the first term of this equation is always smaller than 0. The switching gain k_S can then be derived to meet the criteria of

$$k_S > n \max(|e_\alpha|, |e_\beta|), \tag{12}$$

with n as safety factor, which can be set to 2 [4], [5], [11].

B. Observer Poles

When the current reaches the sliding mode region, the estimated current converges to its actual value. Thus, in the error dynamics the time derivative of the current error becomes zero, i.e. $\dot{\tilde{i}}_\alpha = \dot{\tilde{i}}_\beta = 0$. Then the error dynamic equation reduces to a second order system

$$\begin{aligned}\dot{\tilde{e}}_\alpha &= -\tilde{e}_\beta\,\hat{\omega}_{el} + \frac{1}{L}\,m\,z_\alpha \\ \dot{\tilde{e}}_\beta &= \tilde{e}_\alpha\,\hat{\omega}_{el} + \frac{1}{L}\,m\,z_\beta,\end{aligned} \tag{13}$$

and the estimation error of the back-EMF is

$$\begin{aligned}\tilde{e}_\alpha &= -k_S\,z_\alpha \\ \tilde{e}_\beta &= -k_S\,z_\beta.\end{aligned} \tag{14}$$

Rearranging Eqn. 14 and inserting it into Eqn. 13 yields

$$\begin{aligned}\dot{\tilde{e}}_\alpha &= -\tilde{e}_\beta\,\hat{\omega}_{el} - \frac{1}{L}\,\frac{m}{k_S}\,\tilde{e}_\alpha \\ \dot{\tilde{e}}_\beta &= \tilde{e}_\alpha\,\hat{\omega}_{el} - \frac{1}{L}\,\frac{m}{k_S}\,\tilde{e}_\beta.\end{aligned} \tag{15}$$

From this set of equations the unknown parameter m can be determined. For this purpose, the eigenvalues of the system must be calculated, from which the characteristic equation of a second-order system with conjugate complex poles in the left half-plane results. These poles can be set by the usual means of control theory [4], [5].

C. Saturation Function

An undesirable inherent phenomenon called chattering occurs in the current estimation due to the switching function. By replacing the signum function with a sigmoid, hyperbolic or saturation function, the chattering effect is reduced. In this work, the saturation function is used to increase the thickness of the boundary layer. The saturation function is shown in Fig. 4 with δ as the boundary layer thickness. The value for δ can be determined empirically in the simulation [5], [12].

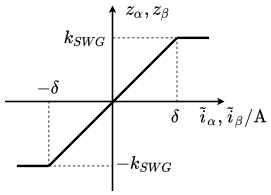


Fig. 4. Saturation function.

V. Speed and Position Estimation

In a traditional way the position can be determined by applying the arctan function to the estimated back-EMF. Calculating the time derivative of the position results in the speed. The drawback of this is that the position and speed estimation is vulnerable to noise. For a more robust estimation a phase-locked loop (PLL) is used. The used structure is shown in Fig. 5 [13].

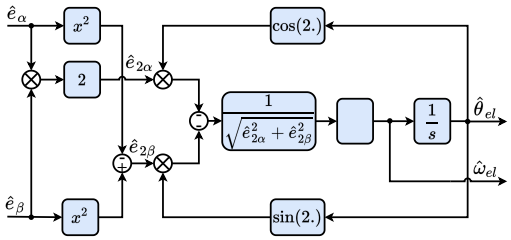


Fig. 5. Block diagram of the PLL.

From the block diagram the estimation error can be determined by

$$\varepsilon = \frac{1}{\sqrt{\hat{e}_{2\alpha}^2 + \hat{e}_{2\beta}^2}}(-\hat{e}_{2\alpha}\cos 2\hat{\theta}_{el} - \hat{e}_{2\beta}\sin 2\hat{\theta}_{el}), \tag{16}$$

where $\hat{e}_{2\alpha}$ and $\hat{e}_{2\beta}$ are modified back-EMF signals defined as $\hat{e}_{2\alpha} = 2\hat{e}_\alpha\hat{e}_\beta$ and $\hat{e}_{2\beta} = \hat{e}_\alpha^2 - \hat{e}_\beta^2$ [13]. Due to the normalization of the modified back-EMF the closed loop transfer function of the PLL simplifies to

$$G_{PLL} = \frac{k_{p,PLL}\, s + k_{i,PLL}}{s^2 + k_{p,PLL}\, s + k_{i,PLL}}. \tag{17}$$

The poles of the system can be set on the real axis by $k_{p,PLL} = 2\rho$ and $k_{i,PLL} = \rho^2$ [5].

VI. Discretization Delay Compensation

Since the control of the system is executed on a microcontroller and the phase currents have to be measured, delays occur due to the discretization. These occurring delays are shown in principle in Fig. 6. There, the duration of the control cycle in which all calculations are performed is 50 μs. In parallel with the execution of the control, the ADC samples the currents sequentially at 200 kHz. At the beginning of the control cycle, these sampled currents are read from the ADC register. The currents read from the register were measured in the previous cycle, so the currents used in the controller are delayed by a maximum of one control cycle. The same applies to the PWM high/low side output. So in the worst case, the discretization delay is about 100 μs or 2 control cycles. This results in a phase shift of the input current that depends on the motor speed. To compensate for this phase shift, the phase shift is simply added to the Clarke-Park transformation of the input currents:

$$\theta_{comp} = n\, T_c\, \hat{\omega}_{el}, \tag{18}$$

where n is an integer for the time delay measured in control cycles, T_c is the period of the control cycle, and $\hat{\omega}_{el}$ is the estimated electrical speed.

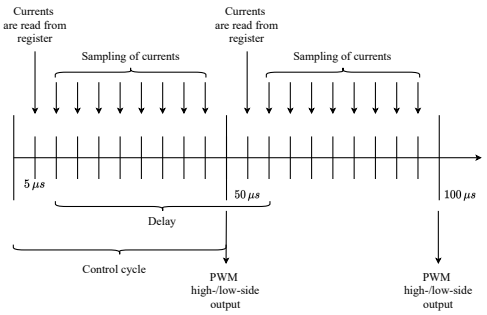


Fig. 6. Basic overview of the timing for data acquisition, output updating and the control execution.

VII. Results

The used motor is an AFPMSM with the parameters listed in Tab. I. The controller parameters are listed in Tab. II, which are calculated with the methods presented in the previous sections. Here is to note that the parameters are designed in the continuous time domain. For the discretization the Euler discretization rule is used.

TABLE I
Motor Parameters

Item	Value	Description
V_{dc}	24 V	DC-link voltage
V	5.56 V	Rated voltage
I_{ph}	15 A	Rated current
f_{out}	1250 Hz	Rated fundamental frequency
$f_{control}$	20 kHz	Control frequency
f_{sw}	200 kHz	Switching frequency
R	12 mΩ	Motor resistance
L	4 μH	Motor inductance
p	5	Pole pairs
k_t	0.0237 $\frac{\mathrm{Nm}}{\mathrm{A}}$	Torque constant, four systems
J	$291.27 \cdot 10^{-6}\,\mathrm{kg\,m^2}$	Moment of inertia

TABLE II
Controller Parameters

Item	Value	Description
$k_{p,c}$	0.01	P gain for current control
$k_{i,c}$	30.15	I gain for current control
$k_{p,s}$	1.22	P gain for speed control
$k_{i,s}$	5.39	I gain for speed control
k_S	16	Switching Gain for SMO
m	1.14	Feedback gain for SMO
δ	80	Boundary layer thickness
$k_{p,PLL}$	200	P gain for PLL
$k_{i,PLL}$	10000	I gain for PLL

A. Simulation

The simulation is performed only for one system out of 4 thus all results are valid for one system but can be transferred to 4 systems. Fig. 7 shows the simulation results for the tracking performance at steady state an no-load expect of a simulated bearing of the observer at a mechanical shaft speed of 10000 rpm. In the first subplot the actual alpha current and the estimated one is depicted, where at this speed a small chattering is present. However, in the other subplots it can be seen that the tracking of the back-EMF, electrical angle and the electrical rotational speed works with the used parameters. The ripple in the current is due to the high switching gain. However in the back-EMF the chattering is not present due to state space extension by the back-EMF and the observer pole design.

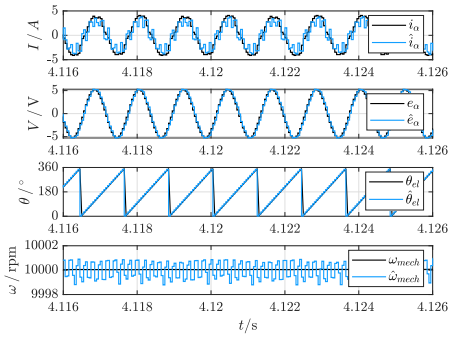


Fig. 7. Simulated steady state observer tracking performance at a mechanical shaft speed of 10000 rpm.

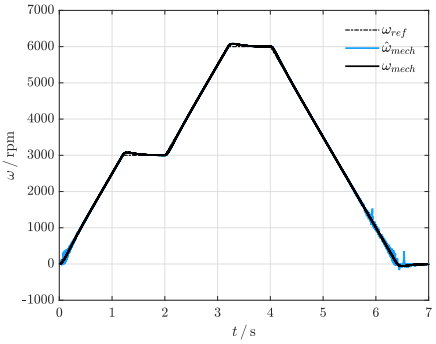


Fig. 8. Speed controller performance with the estimated values.

To validate the speed and current controller pa-

rameters a speed profile is simulated with no-load. This speed profile is shown in Fig. 8. As it can be seen at low speeds the estimated speed is oscillating and also the estimated angle error is quite high. The reason for that is that the system is not observable at zero speed. Furthermore, at low speeds the back-EMF is to small so that the back-EMF estimation error is relatively high. For this, in the simulation a startup is needed where at low speeds the motor dynamics are used to measure the angle and speed and after a certain threshold the system switches to the sensorless control algorithm.

B. Measurement

In this section, the measurement results on the motor are presented for one out of four systems. For the aforementioned startup, an I/f startup method is used up to a speed of 500 rpm. Fig. 9 shows the step response of the shaft speed from a set point of 3000 rpm up to 3300 rpm. Also shown are the estimated speed $\hat{\omega}_{mech}$, the step response calculated from the designed transfer function $\omega_{mech,TF}$, and the speed calculated from the fundamental frequency of the current ω_{mech}. It should be noted that $\omega_{mech,TF}$ is designed for one system and is implemented for four systems in the microcontroller, resulting in a deviation by a factor of 4.

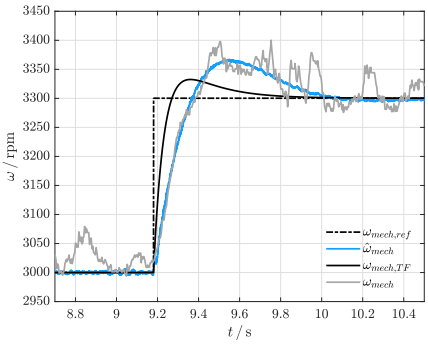


Fig. 9. Step response of the estimated speed, the measured speed and the calculated one from the transfer function.

Fig. 10 shows the tracking performance of the observer as described in the previous section of the simulation. Comparison of simulation and measurement shows that they match, except for the angle, which is in the other direction in the measurement. This is due to the implementation on the microcontroller.

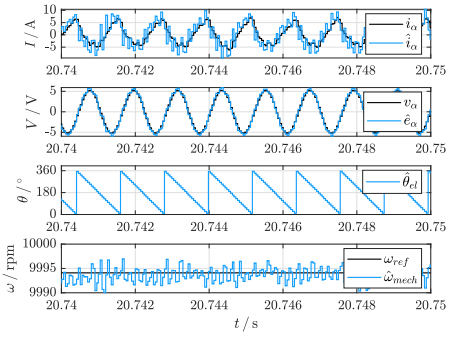


Fig. 10. Measured steady state observer tracking performance at a mechanical shaft speed of 10000 rpm.

The observer performance is quantified in terms of the percent speed error $\varepsilon_{\hat{\omega}}$ in Tab. III. Therefore, the estimated speed is subtracted from the reference speed and divided by the reference speed. This calculation can be considered valid, as it is shown in Fig. 9 that the actual and estimated shaft speed are in the same range.

TABLE III
OBSERVER PERFORMANCE.

$\omega_{mech,ref}$ / rpm	$\varepsilon_{\hat{\omega}}$ / %
2000	±0.30
4000	±0.13
6000	±0.05
8000	±0.06
10000	±0.05
12000	±0.05
14000	±0.03
15000	±0.06

The motor is also tested using 2 systems and the results are as expected. Furthermore, with this imple-

mentation the motor can be accelerated to 15000 rpm in 4 s and 1 system so that the current limits are not exceeded. So, if 4 systems are used the motor is capable of accelerating to 15000 rpm in 1 s.

VIII. Conclusion

It is shown that with the proposed observer and deceleration compensation, the speed of 15000 rpm is achieved. The system shows robust performance over a wide speed range from 500 rpm to 15000 rpm. Moreover, the use of 4 systems works while the full-order position observer is applied to only one system and the angles for the other systems are adopted by knowing the phase shift between them. From the measurements it can also be seen that the motor can be accelerated up to 15000 rpm in 4 s, if one system is in use and the current limits are not exceeded. In the proposed method, only the estimated speed and angle are used in the FOC, this could be extended to the estimated current in further work for more sophisticated delay compensations. For a more robust startup method, especially for startup under load, a more sophisticated startup method, i.e., a high-frequency injection method, can be used.

Acknowledgment

The author would like to thank Maurizio Incurvati, Martin Schiestl and Lukas Wild for providing their expertise and previous work on the system.

References

[1] Zhao, Han, Dai, and Hua, "Study on the Electromagnetic Design and Analysis of Axial Flux Permanent Magnet Synchronous Motors for Electric Vehicles," *Energies*, vol. 12, no. 18, p. 3451, 2019.

[2] B. Hollaus, "Drive Systems: 2nd Semester Master: Reader for the course," 2021, Management Center Innsbruck.

[3] C. J. V. Filho, D. Xiao, R. P. Vieira, and A. Emadi, "Observers for High-Speed Sensorless PMSM Drives: Design Methods, Tuning Challenges and Future Trends," *IEEE Access*, vol. 9, pp. 56 397–56 415, 2021.

[4] Z. Chen, M. Tomita, S. Doki, and S. Okuma, "New adaptive sliding observers for position- and velocity-sensorless controls of brushless DC motors," *IEEE Transactions on Industrial Electronics*, vol. 47, no. 3, pp. 582–591, 2000.

[5] G. Wang, *Position Sensorless Control Techniques for Permanent Magnet Synchronous Machine Drives*. Singapore: Springer Singapore Pte. Limited, 2020.

[6] N. Mohan, *Advanced electric drives: Analysis, control, and modeling using MATLAB/Simulink®*. Hoboken, New Jersey: Wiley, 2014.

[7] M. Incurvati, "Sensors & Actuators: 1st Semester Master: Reader for the course," 2020, Management Center Innsbruck.

[8] A. Mehrle, R. Stärz, and P. Wimmer, "Closed Loop Control: Frequency Domain Methods: Reader for the course," 2016, Management Center Innsbruck.

[9] N. Mohan, *Electric machines and drives: A first course*. Hoboken, NJ: Wiley, 2012.

[10] V. Petro and K. Kyslan, "Design and Simulation of Direct and Indirect Back EMF Sliding Mode Observer for Sensorless Control of PMSM," *Power Electronics and Drives*, vol. 5, no. 1, pp. 215–228, 2020.

[11] T. Furuhashi, S. Sangwongwanich, and S. Okuma, "A position-and-velocity sensorless control for brushless DC motors using an adaptive sliding mode observer," *IEEE Transactions on Industrial Electronics*, vol. 39, no. 2, pp. 89–95, 1992.

[12] C. Gong, Y. Hu, J. Gao, Y. Wang, and L. Yan, "An Improved Delay-Suppressed Sliding-Mode Observer for Sensorless Vector-Controlled PMSM," *IEEE Transactions on Industrial Electronics*, vol. 67, no. 7, pp. 5913–5923, 2020.

[13] H.-K. Hoai, S.-C. Chen, and C.-F. Chang, "Realization of the Neural Fuzzy Controller for the Sensorless PMSM Drive Control System," *Electronics*, vol. 9, no. 9, p. 1371, 2020.

Martin Maier is a Master-student at the department Mechatronics and Smart Technologies with the specialization to electrical engineering at the Management Center Innsbruck.

Evaluation Concept for Analyzing Test Bench Data from Field Failures of High-Voltage Batteries in Passenger Cars

Jan Mohr, Christoph Lörcher (supervisor), and Bernhard Hollaus (supervisor)

***Abstract*—The increasing number of electric vehicles raises the question about an efficient field failure analysis for high-voltage batteries. If one fails after the vehicle being delivered to the customer, it is sent back to the manufacturer for comprehensive failure analysis. To reproduce sporadic faults as well as permanent faults, a load test is used. For high-voltage batteries, this includes a corresponding test bench with temperature control. The results from the test bench should provide the responsible engineer with information about the faulty component. An appropriate system for processing the test bench data is imperative for the analysis of the defective parts. The case of the field failure analysis must be considered separately to have as much information as possible about the faulty component. An implementation of this problem is based on the development method of the V-model for mechatronic systems. Through a survey for stakeholders and the corresponding results, a catalog of requirements for the data evaluation is provided. The survey yields important insights into what exactly stakeholders want, which are included in the presentation of the protocol. The catalog of requirements also serves as a basis for the criteria of the weighted sum model for the selection of a tool. The result of the analysis is that a data evaluation implementation with the Python programming language is the most sensible method. Oriented at the V-model a high level system draft is developed which is broken down in the further process on the individual components. Based on system design and component specification, a fundamental framework for data evaluation is implemented using the Python programming language. This receives the measurement data in a special National Instrument format via the interface to the test bench. As a final result, the evaluation program transfers a structured protocol with the presentation of all relevant results in a document. This helps the responsible engineer to evaluate the measurement data from the test bench and to find the defective component. The Python program serves as a fundamental framework for the evaluation of the results from the load test bench since it has not yet been completely put into operation. Further test steps in the program must be implemented into the evaluation during the respective commissioning of these steps.**

***Index Terms*—High-voltage batteries, Data evaluation, Python, Field failure analysis**

I. Introduction

IN the course of a climate-neutral society, politicians and the automotive industry are increasingly focusing on electric cars. Since electric vehicles do not emit any pollutants like conventional combustion vehicles, this technology should help to achieve the required climate targets. Companies such as the Mercedes-Benz AG, among others, are focusing on vehicles with electric drives. In addition to the electric motor, the core components of an electric vehicle include the high-voltage battery, which stores the electric energy to operate the motor. As with any other component of a technical system, it can always happen that something does not work properly and therefore the defective component must be replaced. At this point, the process of the damaged parts analysis field starts within the Quality Engineering Center of Mercedes-Benz AG. Customers have defective parts replaced within a workshop and the vehicle manufacturer goes in search of the cause of the fault. In the case of purchased parts, the supplier is usually responsible for analyzing the failed components. In the case of high-voltage batteries, this is an internal product of the Mercedes-Benz AG, so the latter is also responsible for the analysis and the fault rectification process. The analysis of field defective components follows the predefined process of the Association of Automobile Manufacturers, which is also followed by the Stuttgart-based group. Thus, within a standard analysis of the defective high-voltage battery, permanent faults are to be detected that indicates the defective component. There are cases in which this standard test is not sufficient and the battery must be subjected to a load test. In the case of the battery, this is done within a special load test bench in which the battery can be tempered so that test sequences can be carried out under extreme conditions. This is currently being commissioned at the Mannheim plant by an external supplier. For this, it is being conceptually worked out how the data from the test bench can be handled. The research question of how and with which program the results from the test bench can best be processed and presented is to be clarified. In the course of the implementation, the point will also be addressed as to how far the evaluation is adapted to the needs of the field failure analysis.In the course of the development, a tool is selected using a user survey and the subsequent weighted sum model. With the resulting tool, a program is created that provides the basic framework for the subsequent evaluation of the data. After the complete commissioning of the test stand, this only has to be edited according to the missing test steps.

J. Mohr study at MCI, e-mail: mj0677@mci4me.at

II. Methods

A. General Approach

To evaluate the test bench data from a load test bench for high-voltage batteries, a system is required that has been developed for this purpose. As a development methodology a logical approach is the V-model for mechatronic systems after the guideline 2206. This model makes it possible to maintain

an overview over the system even of a complex one. The following graphic 1 shows the V-model, adapted to the specific use case of this project.

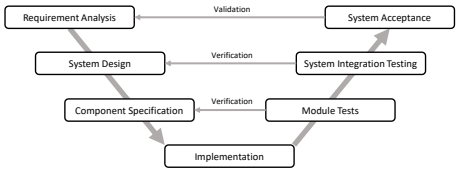


Fig. 1. V-Model as a development method for mechantronic systems modified to the present application.

The starting point is the requirements analysis, in which it is clarified what the program should implement and what the users would like to have in detail. In the system design, the first high level design of the program is sketched. The customer's requirement are further refined in the following steps until they are ready for implementation. After the implementation of the components, these must be verified first, to find possible errors in the program. This procedure is repeated as long as all components of the program are verified. Subsequently, during the system integration, the individual modules are integrated into a system and verified as a whole. It can be seen, therefore, that the V-model not only takes the straight path from start to finish but also allows for iterations to correct any error made in previous stages of design. Finally, the system is released with the validation testing. Thereby, it is checked again whether the entire program corresponds to the requirements. Since the program is a conceptual design, the main focus is on the requirements analysis to meet the needs of the users for a later implementation and to address the research question.

B. Requirement Analysis

To meet the requirements for the end user of the evaluation, a survey is used. Forced by the current circumstances of a global pandemic and the associated working from home, the survey is conducted exclusively online. An online survey mainly reduces the time effort of the respective employees. The survey targets only a small number of employees since the test bench is used exclusively by this group of people for the specific case of field failure analysis within the quality engineering center. Among these stakeholders are on the one hand quality engineers, whose main field of activity is the support of the high-voltage batteries, and on the other hand worker from the same field of activity, who are responsible for the logistic and technical processing of the batteries. The individuals will be contacted individually and personally with a request to complete the online survey. However, to filter out random other participants, they are still asked once again about their work area and function, but these do not play a role in the further results. The following closed and open-ended questions are part of the survey:

1) What is your department?
2) What is your function?
3) How important is a structured presentation of results from the test bench to you?
4) How important is it to you that it is easy to use the tool that is supposed to display the results of the test bench?
5) Assuming all tests were passed: Is it important to you to get more detailed information from the test steps?
6) Is a graphical representation of results helpful in troubleshooting?
7) You have problems displaying the results, do you want a quick solution to the problem or can it take longer?
8) If you have new suggestions for the protocol/evaluation, how would you like to implement them?
9) What else is important to you when presenting test bench results?

The answers to the survey are included in the requirement catalog. The survey shows that all users want a structured presentation of the results. The answers to the next question also show that users prefer the program to be easy to use. On the other hand, the users have a neutral attitude to the question of whether further information should be provided when all tests are passed. Furthermore, most users would like to see a graphical representation of the results, if possible. When it comes to the time it takes to solve problems, users say that it does not have to be done immediately. Nevertheless, they would like to be able to approach a contact person in their team if they have problems or suggestions for improvement. In the last openly formulated question, aspects should come up that have not yet been considered. Once again, it became clear that the users find it important to have a clear and concise presentation of the data and, if possible, to be able to compare the results with others.

The survey results in a catalog of requirements for the data evaluation. This includes the following points:

- Purpose - The data evaluation shall be used to process the data from the load test bench in such a way that a protocol is created which presents the results from the test steps in a structured manner and the test procedure can be identified. This should help to identify the defective component from reproduced fault patterns.
- Target group - The target group of the application are the quality engineers and workers of the quality engineering center at Mercedes-Benz AG, in whose area of responsibility the analysis of the high-voltage batteries lies.
- Functionality - In terms of purpose, the application must output the read-in measurement data in a clear protocol.
- Performance - Only one evaluation will take place at a time on a test run. A load test can last several hours or days, so the evaluation does not have to be available immediately. However, if an evaluation is desired, the application must be working.
- Accuracy - The application must be able to correctly interpret and output all relevant data from the test sequence. Important data must be shown and not lost in the process.
- Reliability - The application must be available after each test run and if problems occur, they must be fixed.
- Security - Since the tested high-voltage batteries are

faulty field returns, their analysis results must also be handled confidentially.

- Maintainability - If errors occur in the application, they must be fixed within the team in a few days to a week from a person within the team.
- Testability - It must be possible to test the function of the application using test data from a golden sample or other batteries.
- Flexibility - As the number of new battery types and generations increases, a high degree of flexibility is important to insure the use of the system in the future.
- Portability - It is possible to provide additional hardware specifically for the load test bench. However, in terms of cost savings, it is advantageous if the application could run on each user's personal computer.
- Interfaces - The software interface of the application is to the test bench, which outputs the measurement data in a TDMS format. The application must be able to process these. The other interface is to the engineer or worker, who will look at the protocol afterward. For easy display and sharing of the data, the output to them must be a PDF document.
- Documentation - Documentation must be available for the use and further development of the application.
- Technical support - The application must be able to run on its own once it is up and running. However, a team of internal experts must be available for queries, problems, or further development.
- Technical requirements - The application must be able to run on a Daimler AG computer without administrator rights or further requirements.
- Costs - In the course of many cost-cutting measures within the company, the focus should be on an application that is free or as inexpensive as possible.

In a pre-selection of tools, one application each was chosen from the categories internal application, external application, and open-source solution. In the internal application category, the Excel template AWP, which was designed in the development department and is already being used on production test benches, is evaluated. As an external application, the software vSignalyzer of the company Vector Informatik GmbH is evaluated, since the Mercedes-Benz AG cooperates strongly with this company. As an open-source solution, an evaluation with the programming language Python in an integrated development environment is evaluated. These three tools are evaluated in the weighted sum model in the following table I.

	Weight	AWP	Value
Presentation of results	5	6	30
Flexibility	3	2	6
In-depth analysis	3	1	3
Maintainability	3	5	15
Technical support	4	10	40
Dependency	5	3	15
Portability	1	4	4
Interfaces	5	2	10
Security	2	9	18
Costs	3	4	12
Sum			**153**

	Weight	vSignalyzer	Value
Presentation of results	5	10	50
Flexibility	3	8	24
In-depth analysis	3	10	30
Maintainability	3	5	15
Technical support	4	3	12
Dependency	5	8	40
Portability	1	6	6
Interfaces	5	10	50
Security	2	10	20
Costs	3	6	18
Sum			**265**
	Weight	Python	Value
Presentation of results	5	10	50
Flexibility	3	10	30
In-depth analysis	3	6	18
Maintainability	3	7	21
Technical support	4	1	4
Dependency	5	9	45
Portability	1	10	10
Interfaces	5	10	50
Security	2	10	20
Costs	3	10	30
Sum			**278**

TABLE I
COMPLETED WEIGHT SUM MODEL FOR SELECTING THE TOOL.

From the weighted sum model, after the evaluation, it is concluded that a solution with the programming language is the most suitable for this application.

C. System Design

Following the V-model, the next step is the system design. The following flow chart 2 shows the high level design of the program to be built.

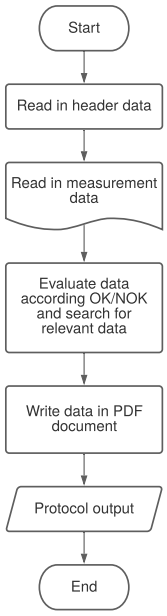


Fig. 2. Rough system design of the evaluation, shown in a flow chart.

First, the header data of the battery are to be read into the program by a graphic surface, since these are not transferred over the measured data. After reading in the measurement data, it is necessary to evaluate for passed and failed tests and thus to select the relevant data for the protocol. Finally, the data is written to a PDF document that is to be output to the user as a protocol.

D. Component Specification

In the previous high level design, the flow of the main program was described. In the next level of detail, the individual components are specified. Based on the result of the weighted sum model and the requirement catalog the evaluation in Python is selected. The view of the high level design shows that the implementation of the individual functions in Python needs different program libraries as Import, which are explained in the following and applied in the implementation. The general information, also called header data, is read by a user input via a GUI. For this, the program library TKINTER is available in Python. In a separate window, a GUI can be created with it, which makes it possible to display forms and provides a template for the input of the header data. These then can be read in and used for further processing. To load a TDMS file into a Python program, another program library is needed. With NPTDMS it is possible to read files in TDMS format and store them in data frames. To access the different groups and channels within the file, it must be iterated through. The program library PANDAS also helps to evaluate the measurement data. It is a tool for the management of data and their analysis, especially for access to numerical tables. The next step from the high level design is to plot the data and write it into a PDF document. For the graphical representation of data in Python, the program library MATPLOTLIB is available. It allows the creation of diagrams and plots in different formats. This is essential for the creation of a protocol. For writing the data and graphics into a PDF document the library PYFPDF is available. It is the basis for the creation and output of PDF documents.

E. Implementation

Based on the requirements profile, the high level design, and the component specification, an implementation with Python is target. The implementation takes place analogously to the V-model. That means that after the implementation of the individual components these are verified and afterward the integration tests take place. The first component to be implemented is the graphical user interface based on TKINTER for reading the header data. The program flow chart of this subprogram is shown in the following graphic 3. The correct input is checked by an internal query for the correct string length and characters so that the user cannot make any mistakes here. The data is transferred to the main program via a button.

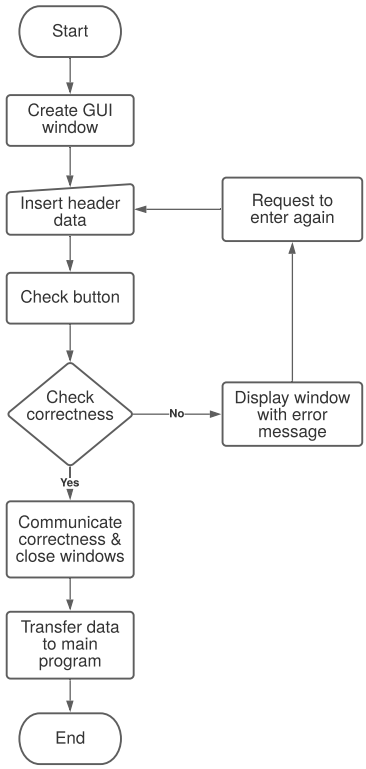


Fig. 3. Program flow of the function for calling up the header data via the graphical user interface.

The main program of the evaluation starts with the call of the header data from the component described before. The flow chart in the graphic 4 shows the superior flow of the data evaluation. First, the measurement data must be fed to the program. To get the path to the measurement data, a GUI opens to select the file. This time TKINTER is not used for this, because it has problems with calling files. Instead, another program library EASYGUI is used for reading the file path. Then the measurement data is read with the help of the NPTDMS library and the measurement properties are stored in a PANDA data frame. This data contains all general information about the device under test, such as the minimum and maximum cell voltages or the number of cells in this battery. From now on, the creation of the PDF document, which will later represent the protocol, already begins. Here the evaluated results are copied step by step. The file name results from the previously filled in header data. Before the first data is written to the document, the PDF document is first defined, such as page size and font. Once this is done, the data that has already been readout, such as the header data and properties, can be written to the document. By the way, these are included in all checks as general information. The writing of the document is implemented with the library

PyFPDF and the methods $cell$ and $multi\_cell$. With these methods, the content of the previously generated data frame is retrieved and written into a cell with a predefined size on the PDF page.

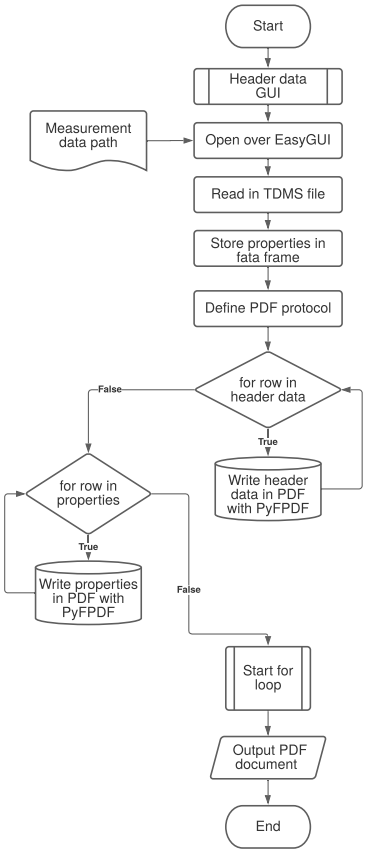


Fig. 4. Program flow of the main program for data evaluation of test bench data.

As can be seen in the previous program flow diagram, the main task of the evaluation is located within a loop sub-function. This has to do with the structural design of the measurement data files, which are divided into higher-level groups and lower-level channels in which the data are located. This sub-function can be seen in the next graphic 5. There it can be seen that by different queries the program falls into different scenarios, depending on the group name and the test result. The test result of the respective test is read from the measurement data and the variable belonging to it is changed accordingly to react to it during the evaluation.

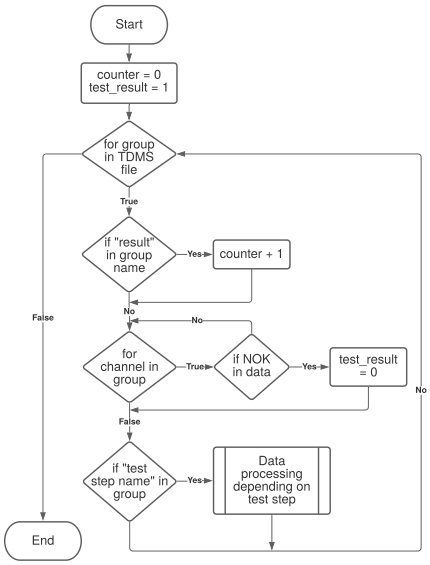


Fig. 5. Program flow diagram for the internal loop for evaluating the data.

Over different inquiries after the group names, the further process of the program is determined in the sub-function of data processing. First for every result group the test step properties are read from the measurement data and stored in the document using $cell$ and $multi\_cell$ methods. Next, the results are read from the measurement data. These are located in the individual channels within a group. Not all channels are relevant for the later output in the protocol, therefore it must be determined before, which information is important and which is not. This is essentially done to ensure that the protocol, as desired in the survey, presents the relevant data in a structured manner and in a scope that is not too large. Even within a group, a loop is used to iterate through the individual channels. In doing so, the channels that are not relevant to the protocol are skipped. This is done within the data processing and ultimately represents the evaluation of the data from the test bench. In the end, this results in a finished PDF document that contains the corresponding evaluation from the test bench data. This is output at the end of the program.

F. Verification

The Python program described in the previous chapter serves as the fundamental framework for the data evaluation of the load test bench and must be edited accordingly for further tests apart from the light profile of the EB200 when they are put into operation. The verification of the individual components and the system serves the purpose of determining whether the results of the respective development step meet the specifications. This is always done step-wise in terms of the V-model and is primarily used for finding errors and for quality assurance. To adhere to the V-model from figure 1, the bottom-up principle is used here. In other words, the individual

components are verified first, followed by system integration, and finally, the system is validated with system acceptance. With the verification of the individual components, which are to be seen graphically again in the diagram 2, there is on the one hand the static procedure and on the other the dynamic testing. With the static one the program is not executed, but only the code is inspected. In the dynamic method, the program is executed and its results are inspected. In the case underlying here, the dynamic method is used, since first test data are already available to test the evaluation. The aim of testing should be to detect as many errors as possible within the tests and to offer the end-user software without any bugs. Since the V-Modell can be seen as a perpetual cycle, the tests described in the following are already partly carried out in parallel with the implementation.

III. Results

A. Protocol

The Python program, described for the evaluation of test bench data from a high-voltage load test bench has as result a protocol. To validate the results, a comparison is made with the requirements of the users and the requirements catalog. In the following, individual excerpts from a protocol with example data are shown to explain the presentation method. The company logo is used as a standard on all Mercedes-Benz AG protocols. To remain within the corporate identity of the company, this can also be found on the test protocol of the load test bench for high-voltage batteries. Below the headline are the header data for the tested battery, which the user has previously entered via the GUI. This part should be clearly distinguished from the rest of the data, as this in itself initially has nothing to do with the test, but is only used for assignment. The following figure 6 shows a section of the protocol header.

Mercedes-Benz

Test Protocol EB200

QEC Reference Number	502100158
Battery ID	555666
A-Part Number	A2053403700
VIN	2052141F999999
Test Bench ID	1
Battery Type	EB200
Employee	Employee 1
Date (Test Date)	01/01/2021

Fig. 6. Excerpt from the protocol on the header data of the battery test.

In the next step, the properties of the battery were added to the program. These are structured from the representation in the same way as the header data. A section of it can be found again in figure 7.

Properties:

Name	IBN_20210602_2_20210602_084557
Min. system voltage [V]	240.000
Max. system voltage [V]	403.200
Min. cell voltage [V]	2.500
Max. cell voltage [V]	4.200
Amount of cells [-]	96

Fig. 7. Excerpt from the protocol on the properties of the battery test.

According to the survey, it is very important to the users that the data is presented in a structured way in the protocol. To reduce the information to the essential, the descriptive statistical method of a presentation in tabular form is used. Thus, in the section in figure 8, it can be seen which steps and results are involved. Likewise, a clear demarcation was required to see whether a test was passed or not. Using a quality contrast and the color green, which generally stands for okay, the result of the test step is once again highlighted. If a test step is not passed, the result is highlighted in red.

Step: Start

Parameter Description	Parameter Value
Max. duration	90
Set whitelist fault memory global [-]	
Set voltage LV [V]	13.8
Set max. current LV [A]	3

Step	Actual Value	Unit	Set Value	Set Value LL	Set Value UP	Result
Result step "Start"	0		0	0	0	OK
Step time	7.9	s	7.9	7.9	7.9	No findings
Step result: OK						

Fig. 8. Excerpt from the protocol on the one test step of the battery test.

Here, a test step with a positive result is shown as an example for the protocol. The final protocol is composed of the parts described above and the results from further test steps. As described in the program flow, the display of the test step is reacted to differently depending on the result of the test. If, for example, a test step turned out negative, graphical representations of voltage and temperature curves are used here.

B. Validation

In the implementation of the program, the verification of the software was carried out continuously according to the V-model to see if the program is implemented correctly. The validation of the program must take place afterward and gives information on whether the program fulfills its purpose or not. For this purpose, the needs of the customer are considered, as well as the catalog of requirements and the applicability of the program. The applicability can be tested only by the users themselves. However, since the test bench has not yet been put into operation, there is currently no acute need for this. Nevertheless, it is made as simple as possible for the user to operate the evaluation by the application of a GUI and the simple execution of the program without further actions. The general purpose of the program, according to the requirements catalog, is to output a protocol that presents the results from the test steps of the load test bench in a clear and structured manner. By looking at the protocol in the results, it can be seen that the program fulfills this purpose for the target group. Most of the customer needs have already

been matched within the presentation of the results, such as the structural and graphical representation of the protocol. With the selection of Python for implementation, out of the weighted sum model, the program can also remained within the team. Thus, there is direct access to the person in charge.

With enough prior Python knowledge and the documentation on the program, users can also make changes to the analysis in consultation with the team. A further look at the requirements catalog should confirm whether the other requirements have also been met. Most of these requirements, such as the cost, portability, flexibility, system requirements, and security have already been met by the selection of the programming environment with Python. The resources for queries and maintenance of the system must still be provided by the department. Furthermore, the program serves the input interface with the reading of the measurement data files and the output interface with the output of a PDF protocol. This also fulfills the requirements. Repeated testing with different data sets from the load test bench shows that the program also repeatedly outputs what is required of it. Thus, it can be said that the program provides the required functionality and is classified as valid according to the ISO 9000 series of standards.

IV. Conclusion

The load test bench for high-voltage batteries, which is currently being commissioned at the Mannheim plant, runs through various test sequences to reproduce sporadic faults in high-voltage batteries. This is intended to identify the faulty component within the battery. The output of the test bench is a series of measurement data. The research question is, what is the best way to process and present the results from the test bench. To get to the bottom of this question, a survey has also been created in the spirit of the sub-hypothesis. This should give the direction for the elaboration to meet the special needs of the quality engineering center. Using the weighted sum model, the best tool for implementation was selected and finally, an evaluation was implemented with it. As a result of this, the user receives a structured protocol that reflects the results of the load test.

Requirements from the employee survey were "*that it should be clear which tests were passed and which were not*" and that "*the relevant results should include the minimum and maximum values allowed*". Thus, in the output of the protocol, attention is paid to the partial hypothesis that this should be adapted to the field of field failure analysis. These expectations are met with the result. Concerning the research question, it turned out that the best possibility for the representation of the data is an implementation with the programming language Python. With the implementation of the analysis and a structured protocol as output, this expectation was also fulfilled.

The implementation of data evaluation with Python offers users in the quality engineering center much more than just the simple display of results. Here, the user's wishes can be individually catered for. This can be done, for example, with an additional display of the limits from different test steps. Viewed across the entire company, it naturally depends on the use case which evaluation method is the best. The classification of this solution is designed especially for the use case of field failure analysis. However, also possibilities for other areas can be derived from this, because Python has many possibilities for displaying results, which can be adapted very individually in each case.

The data evaluation is currently still limited by the low progress in the commissioning of the load test bench. The fact that this has a rather high time delay, is noticeable in the implementation of the different tests. The fundamental framework of the evaluation, as well as the first tests, have already been integrated. To display the results of further and more extensive test sequences with the evaluation by Python, these must first be implemented in the evaluation. With several embedded test steps in the system, it can be adapted to the user requirements in a different way by performing another round of tuning after the overall implementation. Thus, the evaluation can be further optimized.

An implementation with Python has many advantages, but also involves work when considering the technical support. There must be a responsible person in the team who supervises the data evaluation and expands the system. For this, in addition to the knowledge of the programming language, above all, there must be the necessary manpower. The lack of manpower of the team in the current situation could mean difficulty in the further operation of the evaluation. Furthermore, the evaluation is currently designed in such a way that the user receives a structured protocol that provides him with the relevant information from the test run and leaves him room for interpretation. However, if more in-depth analyses of the measurement data are required, this is possible with Python, but in this case, the effort is considerable. Thus, for this case, it is always necessary to create a separate environment in Python or to fall back on another tool that is specialized for a deeper analysis. Nevertheless, according to the requirements of the users, the evaluation with Python is the best for this use case. With the current results, there is still room for improvement. One possible improvement would be to integrate the evaluation directly into the test bench so that the step of reading the files into the program is automated. As well as the complete process of the evaluation. Here it would be advantageous if this were to start automatically directly after the load test is done. With the integration of further test steps and the execution of further tests, conclusions can also be drawn about previous results, as was the wish of a user during the survey. To do so, a certain amount of basic data must first be available to take this further step. If enough data sets are collected from the testing processes, it is conceivable that neural networks can be used to conclude certain fault patterns. With sufficient data, it is possible to continue research in this direction.

Anyhow the solution with Python is primarily only an isolated solution. Although this is the best solution for the area in the current situation, Markus Schäfer, a member of the board of management at Daimler AG, is also in favor of standardization within the company. Thus, standardized solutions for the load test bench in the quality engineering center are also conceivable. First approaches for standardized evaluations of test benches in the company already exist through the internal system UNIPAS. This is already used in some test benches in production but requires some hardware changes to the test bench itself for implementation, which the current test bench in the quality engineering center does not have. For the acquisition of the next test bench, which is currently in the brainstorming phase, the inclusion of this system would be worth considering and should have to be checked for its possibilities. Another system that is currently under discussion in the area of data evaluation is the MORPHEE software from FEV Software and Testing Solution GmbH. However, before committing to a standard here, a decision must first be made on this topic at the management level.

ACKNOWLEDGMENT

The author would like to thank his supervisor Christoph Lörcher, who supported and motivated him during the making of this master thesis. Special thanks go to the MCI in Innsbruck for the basics learned from the master's program, as well as to the university lecturer Bernhard Hollaus for the supervision on the university side. Finally, the author would like to thank the Mercedes-Benz AG, which made it possible to work on this topic as part of a thesis for the Master's degree and provided the necessary resources.

Jan Mohr is a student in the Mechatronics Department at the MCI Innsbruck/Austria. He started his master's program there in October 2019 and will finish in 2021.

Non-invasive wide-bandwidth current sensor for wide-bandgap devices

Stefan Moser, and Maurizio Incurvati (supervisor)

***Abstract*—Modern power electronics devices like wide-bandgap semiconductors allow for higher switching frequencies compared to existing silicon devices. Measuring the switching currents in those systems requires a wide-bandwidth current measurement sensor. However, available measurement systems require modifications of the PCB due to their size and introduce therefore parasitics into the system or are limited in bandwidth. This paper deals with the design and development of an Rogowski coil based current sensor, which allows for a low parasitic measurement of the switching current in an semiconductor device. The sensor is mounted on an adapter board, which is placed between the component and the PCB. Shielding and the geometry of the windings provide good suppression of interfering adjacent currents. Together with an wideband integrator, fast transients in the nanosecond range can thus be measured. Experimental verification shows good agreement with the simulation. In future applications, the system can be implemented in integrated solutions allowing an on-chip measurement.**

***Index Terms*—GaN, WBG, Current Sensor, Rogowski Coil**

I. Introduction

CURRENT is one of the most important measurement quantities in power electronics. For this reason, some form of current measurement is implemented in almost every power electronic circuit. The measured current is used for control, monitoring and protection purposes. Besides this applications, it is also used in the laboratory to determine switching losses and efficiency.

In Fig. 1 a three-phase inverter is shown as an typical example for an power electronic circuit. Ten different current measurement points are available. One measures the current out of the DC link supply. Three of them, I_U to I_W measure the load currents. The other six, I_1 to I_6, measure the switching currents.

S. Moser and M. Incurvati are with the Emerging Applications Lab, e-mail: stefan.moser@mci.edu.

The requirements for the measurement of the load currents are lower, since the current can be assumed to be continuous due to the load inductance or filtering. It is often in the form of a sine wave and the period duration is usually in the millisecond range, therefore a sensor capable of a bandwidth of some kilohertz is suitable. As this wires are often fed to another component like a motor, no limitations in size of the sensor are given.

The switching currents are in the form of a trapezoid, with two steep switching edges in the order of nanoseconds and a on-time of some micro seconds. Due to the steep switching edges, the measurement of those is demanding and requires a high bandwidth with DC capability. To satisfy this, different sensing technologies are linked.

The use of current sensors has so far always entailed a change to the PCB, therefore changing the properties of the PCB in terms of parasitic inductance. Since these measurements are no longer needed in later operation, it is appropriate in terms of manufacturing and cost optimization if the measurement system can be removed at the end of the prototype phase without causing serious changes to the circuit characteristics.

To achieve this goal, a current sensor based on a Rogowski coil is designed. The transducer is made in such a way that it can be placed as an adapter between the footprint on the PCB and the component

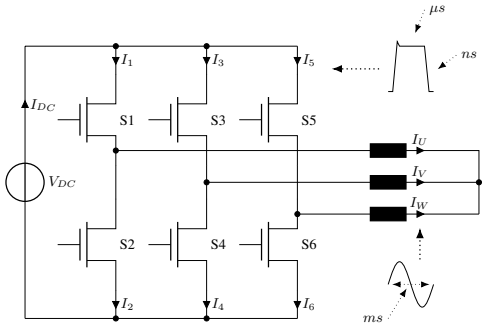


Fig. 1: Three-phase inverter with typical signal characteristics of the current at the measuring points I_5 and I_W.

itself. The signal induced in the Rogowski coil is fed to an integrator. The integration allows the original current waveform to be restored. The measurement system is designed in such a way that an adaptation for DC capability can be made later.

II. Sensing Methods

A. Shunt

One of the most popular and easiest ways to measure current is the shunt resistor method. In this approach, the measurement circuit is interrupted and a resistor is inserted. A voltage drop is caused by the current flow, which is linear according to Ohm's law.

$$U = R_{Shunt} \cdot I \tag{1}$$

This inserted resistor generates losses in the system. In non-ground referenced measurements, additional circuits are required. For small currents to be measured, the resistance value must be adapted accordingly.

Coaxial shunts are used for measuring higher frequency currents. The design of these shunts is chosen in such a way, that the currents are distributed symmetrically and the area spanned by the conductors is minimized. Thus a high bandwidth of over $400\,\mathrm{MHz}$ can be achieved. Downside is the big size of the resistors and the complexity of mounting them on the PCB.However, shunt resistors are not preferred in power electronics, since losses shall be minimized. [1][2]

B. Hall effect

Hall effect sensors generate a measurable voltage, which is proportional to the magnetic field and the bias current. The bandwidth is typically limited to frequencies below $500\,\mathrm{kHz}$ [3]. [2] The Lorentz force acts on a particle with the electric charge q moving with the velocity v in the electromagnetic field. For a given bias current I, the Hall sensor measures the magnetic field value, in which the sensor is situated. This field value represents the current in the area. [4] To increase the sensitivity and selectivity, the sensor is often mounted in the air gap of a magnetic core, since it is sensitive to external magnetic fields and electric currents. [1] The resolution is limited by offset errors, due to the remanence of the field in the magnetic core and noise, due to the small voltages generated. [1][4]

C. Fluxgate

Fluxgate sensor measures current under the influence of magnetic field. Due to its construction, it ensures galvanic isolation. It consists of two ferromagnetic cores, which are periodically brought into and out of saturation by excitation windings. Due to the anti-parallel positioning of the cores, both voltages V_1 and V_2 cancel each other and V_{sense} remains zero. When an external magnetic field is applied, the induced voltages are imbalanced, since one of the cores saturates faster than the other, leading do a rectangular signal with a certain pulsewidth. By demodulating and integrating the sum of the voltages, the signal waveform of the current to be measured can be reproduced.

High sensitivity can be achieved, but the bandwidth is limited to small values below $75\,\mathrm{kHz}$ due to saturation and signal processing. [2]

D. Current Transformer

Current transformer are popular devices to measure current, as their build-up is simple and robust, they do not require external power, have high galvanic insulation and provide a long lifetime while being cheap.

Depending on the structure, it consists of only one or multiple primary windings, as shown in Fig. 2a. On the secondary side, the current on the primary side is reduced with the inversely proportional ration of primary to secondary winding (Eq. 2).

$$I_2 = \frac{N1}{N2} \cdot I_1 \tag{2}$$

The secondary side has either to be shortened or the measurement device has to be mounted, since otherwise high voltages can appear on the secondary side. Depending on the frequency of the measured current, the optimum number of turns differs. The magnetizing current, which is inversely proportional to the frequency requires a high permeable core material and an increasing number of turns. At higher frequencies, parasitic capacitance is the dominant source of errors. As the capacitance rises with the number of turns, an increasing number of latter would increase the error. [1]

E. Rogowski coil

A Rogowski coil is an air coil, which is wound around the current conductor, on which measurements are to be performed. The mutual inductance between the primary and secondary side determines the operating principle, where the output voltage is proportional to the derivative of the current.

$$V(t) = M \cdot \frac{dI}{dt} \tag{3}$$

Together with an integrator circuit, the ac current waveform can be obtained.

$$V_{I_{meas}}(t) = \int V(t) = \int M \cdot \frac{dI}{dt} = M \cdot I \tag{4}$$

Since the coupling M is weak, a return wire loop in the opposing direction is needed, to minimize the impact of fields induced by nearby inductors. The return wire can be a simple wire or can consists of another winding, which must be wound in opposing directions. Latter winding scheme leads to a doubled induced voltage. [5]

As no ferromagnetic materials are used, linearity and a large dynamic range is given. The self inductance and parasitic capacitance form a resonant circuit which limit the bandwidth. [1] For long term measurements the drift of the integrator circuit and noise is a limiting factor. An illustration is given in Fig. 2b.

The performance characteristics are favorable when compared to conventional current transformer, due to the high measurement accuracy and the wide operating current range, as no saturable iron core is present.[5]

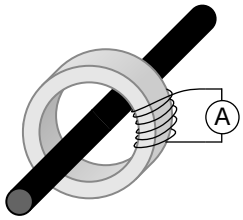
III. Design of Sensor

A. Rogowski coil

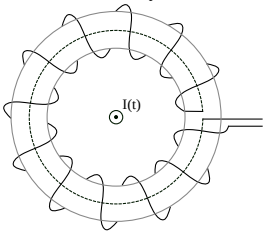
Unlike commercially available coils, which are placed around the current-carrying cable to be measured, this coil is placed in a PCB due to the dimensions of the package. The medium-voltage switches from Infineon used on this board have a width in the order of the diameter of smaller commercially available coils.

In Fig. 3 the lumped-element model of a coil is given. This shows which parameters can be adjusted in principle in the design. This is the inductance L_{par} of the primary side, where the signal path of the power circuit is located. This path has an resistance R_{SR}. The coil L_{coil} is coupled to the primary side via the coupling factor M, which indicates how good the primary side with the secondary side is coupled.

R_{ser} is the resistance of the traces along the coil. C_{par} is the inter-winding capacitance together with the capacitance to the ground of the system. The damping resistor R_{damp} is not part of the coil itself, but is necessary at the input of the system.



(a) Current Transformer with a single primary turn (conductor) and five secondary turns.



(b) Rogowski coil with the conductor in the center, the windings on the air core and the dashed return wire.

Fig. 2: Illustration of two current measurement principles, providing different physical output quantities.

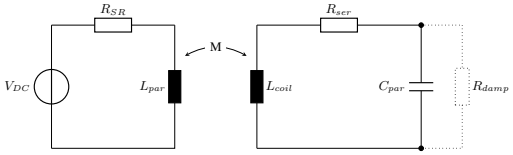


Fig. 3: Lumped model of a rogowski current sensor

When designing the coil, care has to be taken to find a trade-off between the required system bandwidth and the coupling factor. Therefore increasing the number of turns would lead to a higher coupling factor, but lowers the bandwidth, since the inductance of the secondary side rises.

Besides the design of the winding structure itself, care has to be taken to surrounding signals, which have to be prevented from coupling to the system. Besides that, high voltage differentials, which appear

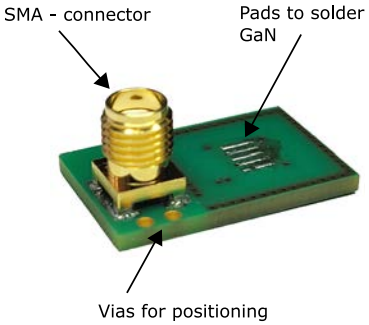


Fig. 4: Adapter PCB with the SMA connector for the measurement output and the solder pads for the GaN-switch.

in switches, can be coupled via parasitic capacitance. Since the coil is not build differential, capacitive coupling is not canceled out, leading thus to measurement errors.

In the design, a four-layer approach is chosen, where the windings are placed on the internal layers. On the external layers, the source-potential is placed symmetrically to act as a shield to the surrounding. Internally, a second internal screen prevents the earlier mentioned capacitive coupling.

The produced prototype of the adapter PCB and the integrator can be seen in Fig. 4.

B. Integrator

1) Methods:

a) Passive Integration: One simple method to integrate signals is to use a passive integrator consisting of an resistor and a capacitor. Integration takes place only for frequencies well above the cutoff frequency.

The limiting factor of using this kind of integrator is that higher frequency signals have a very low magnitude, when using this integrator for wide bandwidth applications. Limiting factor in the design itself is the availability of passive components. [6]

b) Inverting amplifier: To overcome the issue with amplification, an active integration circuit is preferred. The simplest circuit is shown in Fig. 5. The output voltage is described by following equation:

$$v_{out} = -\frac{1}{RC}\int_0^t v_{in}dt + v(0) \tag{5}$$

The feedback resistor R_2 limits the amplification for DC values, reducing amplification of noise.

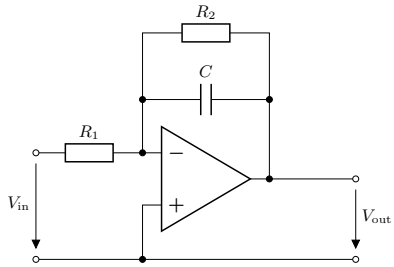


Fig. 5: Inverting op-amp integrator with feedback resistor.

This structure has a disadvantage when measuring fast transients. Due to the fast rising current, a voltage step appears at the input, which will be passed trough the capacitor due to the finite response time of the operational amplifier. The result is a voltage peak in the opposite direction, a so called pre-shoot. [7][8]

c) Non-inverting integrator: Better results provides a non-inverting operational amplifier circuit, shown in Fig. 6, with the output voltage being, if $R_1C_1 = R_2C_2 = RC$ is full filled:

$$v_{out}(t) = \frac{1}{RC}\int_0^t v_{in}(t)dt + v(0) \tag{6}$$

This shows, that the output voltage shows the same behavior as the inverting integrator, but with opposite sign. Due to the passive filter structure in the input, the pre-shoot effect cannot appear, since the capacitor C_1 is charged first.

2) Description: The principle circuit used, is shown in Fig. 7. A non-inverting integrator (1) is

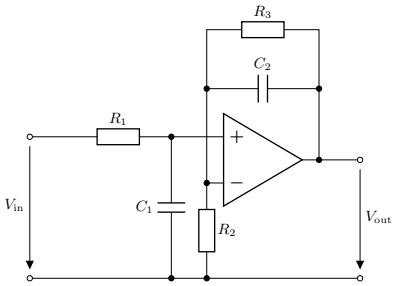


Fig. 6: Non-inverting op-amp integrator with feedback resistor.

implemented with the operational amplifier U_1. Additionally U_1 acts as a subtracting amplifier, subtracting the low-passed input voltage from the feedback voltage. R_4 helps to decouple the load capacitance from the output.

(2) is a non-inverting amplifier, where the amplification is set via R_7 and R_8. (3) amplifies the feedback signal. As the feedback has to be subtracted from the input, an inverting amplifier structure is used. R_{12} and R_{10} set the gain. R_11 is used to compensate the input bias current. The feedback is low-pass filtered in (4) before fed back via the coupling resistor R_{14}. Together with the capacitor of the input filter C_1, the time constant for the feedback can be set.

The produced prototype is depicted in Fig. 8.

IV. Results

A. Measurement setup

To test the sensor, fast rising currents have to be provided. An evaluation board, featuring a half bridge with GaN switches is therefore designed. A so-called double pulse test can then be performed, where the circuit is connected as shown in Fig. 9.

This allows to switch on and off high currents, which imitates the behavior in the applications.

Since the power supply has to provide high power in short time, additional capacitors are placed to keep

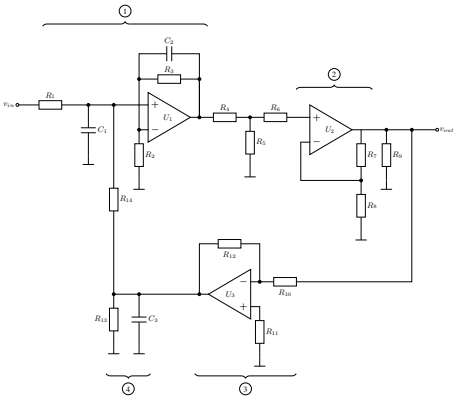


Fig. 7: Principle circuit used for the integrator.

Fig. 8: Prototype of the integrator circuit.

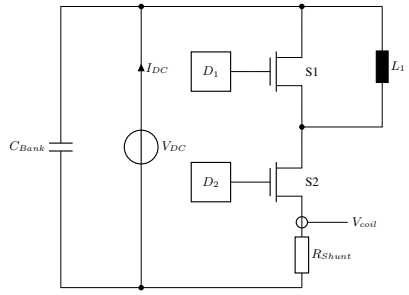


Fig. 9: Principle circuit to conduct the double pulse test circuit.

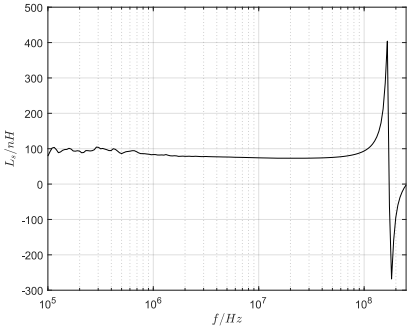


Fig. 10: Series inductance of secondary side of the Rogowski coil.

the bus voltage at the desired value. The capacitors are located near to the switches to minimize the inductance in the power loop.

The signals for the drivers are generated by the XMC4400 drive card, where different on-times can be set. The measured signals are compared to a reference current sensor, which is a coaxial shunt resistor with a bandwidth of 400 MHz.

B. Measurement Results

Fig. 10 shows the series inductance of the secondary side of the Rogowski coil. The measured values show good agreement with the simulated values from Ansys. At 150 MHz, a series resonance can be found.

The mutual inductance between the primary and secondary side of the Rogowski sensor is depicted in Fig. 11. It can be seen that the mutual inductance has a nonlinear behavior due to a resonance effect at 43 MHz. Thus this limits the bandwidth of the sensor.

The switching currents in a double-pulse measurement are depicted in Fig. 12, where the blue colored waveform shows the reference current measurement, measured with the coaxial shunt resistor, and the blue colored waveform shows the output of the proposed

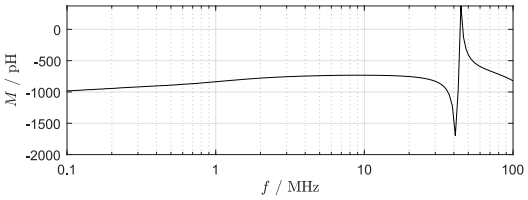
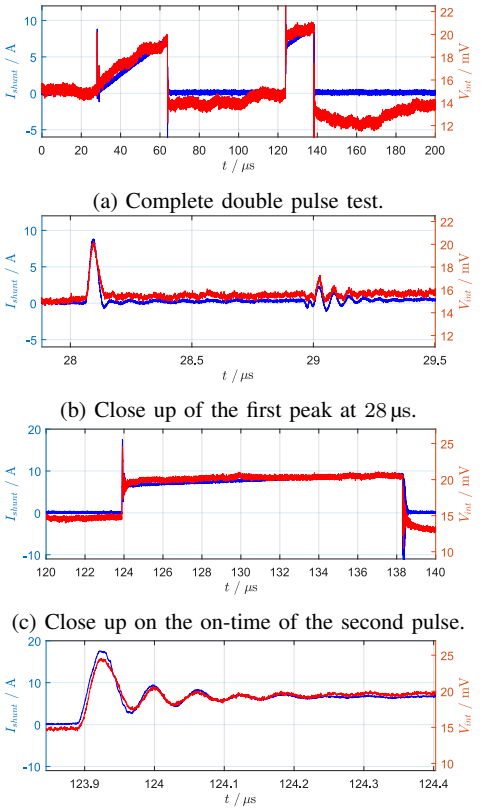


Fig. 11: Mutual inductance measured with the VNA.

sensor. The dc link voltage is selected 24 V. During this measurement, an additional capacitor was placed in parallel to the high side switch, to lower the oscillation frequency. This allows a characterization of the sensor below the nonlinearities in the inductance. The measurement result of the complete result is shown in Fig. 12a, where despite the noisy output signal a good matching with the current waveform can be seen. When the first peak is shown in more detail in Fig. 12b, a good matching of both current measurements can be seen. Also the smaller oscillations at 29 µs are captured. In Fig. 12c, the second pulse is plotted in a close-up. It can be seen, that the pulse is covered in a good matching result for the whole length. However, after the transient is occurring, the output of the sensor has not an sharp change as the reference current measurement shows. In Fig. 12d, the first oscillations of the pulse can be seen, which show good matching. Since the scaling of the figures 12 (b) and (c) is the same, a small offset voltage appears at the beginning. Considering this, both covered signals show the same amplitude.

In Fig. 13, the frequency characteristics of the system are plotted. The integrating behavior of the magnitude shows a $-20\,\mathrm{dB}$ behavior up to $160\,\mathrm{MHz}$. However the phase shift drops earlier, having already $-100°$ at $35\,\mathrm{MHz}$, which limits the bandwidth of the integrator.



(a) Complete double pulse test.

(b) Close up of the first peak at 28 µs.

(c) Close up on the on-time of the second pulse.

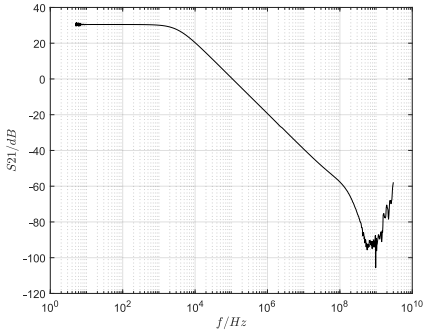
(d) Close up on the oscillations

Fig. 12: Comparison between shunt measurement and the developed sensor at a double pulse test. A additional capacitor with 10 nF is placed in parallel to the high side switch to shift the switching frequency.

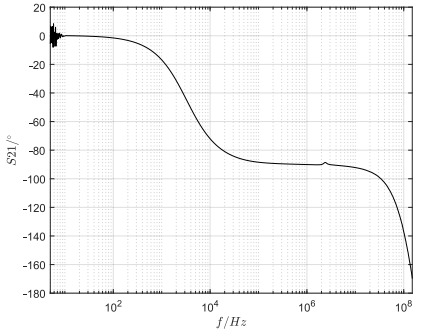
V. Conclusion

This paper analyzes an PCB integrated Rogowski coil to measure the switching current into an GaN-switch. The basic principle in forming a Rogowski current sensor are described. Different aspects in the design of the coil are shown. Limiting factors, which have to be considered during the design are listed. Properties of different integration methods are shown, from which the most suitable variant is chosen. The structure of the integration circuit is shown, which is necessary to handle the small voltages.

First tests of the prototype have been discussed.



(a) Transmission characteristics of the system.



(b) Phase response of the forward transmission.

Fig. 13: Frequency characteristics of the integrator circuit.

Experimental verification shows good behavior of the tests for frequencies below 30 MHz. In future designs, the inductance of the secondary side of the Rogowski coil has to be adapted, so that resonance effects due to parasitic capacitances appear at higher frequencies, thus allowing a wider bandwidth. Additionally, a better measurement setup for higher frequencies, like placing the components on a common metal surface can improve the bandwidth. Digital integration can also increase bandwidth.

REFERENCES

[1] Eren, Halit, Webster, and J. G, "Measurement, instrumentation, and sensors handbook, second edition."

[2] T. Funk and B. Wicht, *Integrated Wide-Bandwidth Current Sensing*, 1st ed., ser. Springer eBook Collection. Cham: Springer International Publishing and Imprint Springer, 2020.

[3] J. Jiang and K. A. A. Makinwa, "A multi-path cmos hall sensor with integrated ripple reduction loops," *Proceedings of technical papers : November 9-11, 2015, Xiamen International Conference Center (XICC), Xiamen, Fujian, China*, 2015. [Online]. Available: http://ieeexplore.ieee.org/servlet/opac?punumber=7378179

[4] S. Tumański, *Handbook of magnetic measurements*, online-ausg ed., ser. Series in sensors. Boca Raton: Taylor & Fracis, 2011. [Online]. Available: http://lib.myilibrary.com/detail.asp?id=325744

[5] A. K. Ljubomir, R. Beresh, M. T. Bishop, R. Javora, B. Magruder, P. McLaren, B. Mugalian, and A. Offner, "Practical aspects of rogowski coil applications to relaying."

[6] Y. Shi, Z. Xin, P. C. Loh, and F. Blaabjerg, "A review of traditional helical to recent miniaturized printed circuit board rogowski coils for power-electronic applications," *IEEE Transactions on Power Electronics*, vol. 35, no. 11, pp. 12 207–12 222, 2020.

[7] W. F. Ray and R. M. Davis, "High frequency improvements in wide bandwidth rogowski current transducers," Lausanne.

[8] Lei Ming, Zhen Xin, Changqing Yin, Manxin Chen, and Poh Chiang Loh, *ECCE 2019: IEEE Energy Conversion Congress & Expo : Baltimore, MD, Sept. 29-Oct. 3*. Piscataway, NJ: IEEE, 2019. [Online]. Available: https://ieeexplore.ieee.org/servlet/opac?punumber=8897530

Stefan Moser is working in the Infineon Emgering Applications Lab in Innsbruck/Austria.

Entwicklung einer Ausbildungsplattform zum modellbasierten Entwicklungsprozess von Automatisierungslösungen am Beispiel von Industrieroboterkinematiken

Julian Muigg und Benjamin Massow (Betreuer)

***Kurzfassung*—Handelsübliche Industrieroboter sind häufig für anspruchsvolle industrielle Anwendungen ungeeignet, weshalb spezialisierte Eigenentwicklungen notwendig sind. Dafür benötigen Ingenieure profundes Wissen über Robotik, Automatisierungs- und Antriebstechnik, welches mit anwendungsnahen Ausbildungsmöglichkeiten vermittelt werden soll. Ziel dieser Arbeit ist die modellbasierte Entwicklung eines Software Frameworks für die Steuerung und Bedienung einer Ausbildungsplattform für serielle und parallele Roboterkinematiken. Nach der Definition der Anforderungen und der Ausarbeitung eines Systementwurfs folgt die modellbasierte Entwicklung und Integration des Software Frameworks mit der *X in the Loop* (XiL) Methode. Die Anwendungsfälle des Delta360 und SCARA Roboters zum Testen des resultierende Software Frameworks zeigen die virtuelle Systemsimulation, die Integration und Verwendung von erarbeiteten Lösungen der Kinematik und Trajektorienplanung sowie die Steuerung der Ausbildungshardware. Die Bedienoberfläche ermöglicht eine manuelle Steuerung des Roboters sowie die Kalibrierung und Abarbeitung einer virtuellen Pick & Place (PP) Aufgabenstellung.**

***Schlagwörter*—Ausbildungsplattform, Entwicklungsprozess, System, Modell, Modellbasierte System-Entwicklung, *X in the Loop*, Virtuelle Inbetriebnahme, Digitaler Zwilling, Industrieroboterkinematik, Automatisierungstechnik, Robotik, SCARA, Delta360**

I. Einleitung

FÜR industrielle Anwendungen gibt es am Markt verschiedene Modelle von Industrieroboter. Häufig sind diese Industrieroboter für anspruchsvolle Anwendungen sowie komplexe Produktionsanlagen ungeeignet und können für die Lösung der Aufgabenstellung nicht verwendet werden. Dafür werden spezialisierte Robotersystem benötigt, welche durch Eigenentwicklung erarbeitet und hergestellt werden. Für die Durchführung dieser Entwicklungen benötigen Ingenieure profundes Wissen über Robotik sowie Automatisierungs- und Antriebstechnik. Themengebiete wie direkte und indirekte Kinematik, Trajektorienplanung, Dynamik und Applikationssteuerung von Robotersystemen sind anspruchsvoll und mit rein theoretischen Ausbildungen schwer zu vermitteln. Deshalb werden anwendungsnahe Ausbildungsmöglichkeiten mit Laborhardware benötigt.

J. Muigg ist Student am MCI für Mechatronik & Smart Technologies, Innsbruck, Österreich, E-Mail: mj4468@mci4me.at.

Benjamin Massow arbeitet am Studiengang Mechatronik, MCI, Innsbruck, Österreich, E-Mail: benjamin.massow@mci.edu.

Manuskript eingereicht am 28. September 2021; revidiert am 28. September 2021.

Damit die notwendigen Kenntnisse der zuvor erwähnten Bereiche anwendungsnah übermittelt werden können, befasst sich diese Arbeit mit der Entwicklung einer Ausbildungsplattform für serielle und parallel Roboterkinematiken. Dabei liegt der Fokus auf der Entwicklung eines Software Frameworks für die Ansteuerung des Robotersystems sowie der Bedienung der Ausbildungsplattform. Des Weiteren soll diese Steuerungsapplikation ein Robotermodell mittels virtueller Systemsimulation und die Laborhardware über eine Programmable Logic Controller (PLC) steuern können. Das Software Framework soll die Möglichkeit bieten, erarbeitete Lösungen zu den zuvor erwähnten Themen sowohl virtuell als auch real zu integrieren und zu testen. Ebenso soll diese Ausbildungsplattform für Demonstrationszwecke ohne zusätzlichen Aufwand vollständig funktionsfähig sein und als Beispiel einer industriellen Anwendung dienen.

Der Aufbau dieser Arbeit ist in mehrere Phasen unterteilt, beginnend mit der Definitionsphase. In dieser werden die Systemanforderungen definiert und analysiert. Anschließend wird in der Konzeptphase ein Systementwurf mit Hardware- und Softwarekomponenten ausgearbeitet. Darauf folgt die Entwicklungs- und Integrationsphase, welche mit Hilfe der virtuellen Inbetriebnahme (VIBN) und der X in the Loop (XiL) Methode durchgeführt wird. Des Weiteren wird für die Ausbildungsplattform eine Human-Machine Interface (HMI) Applikation für die Anlagenbedienung entwickelt sowie Testapplikation für die Validierung und Verifizierung des Systemmodells. Die resultierende Steuerungsapplikation wird schlussendlich in eine *Beckhoff* Steuerung integriert und damit die reale Anlage gesteuert.

II. Definitionsphase

Die Ausbildungsplattform dient als anwendungsnahe Lernplattform für Ingenieure. Diese Ausbildungsplattform besteht aus einem SCARA Roboter, als Vertreter der seriellen Kinematik, und einem Delta360 Roboter, als Vertreter der parallelen Kinematik. Die Abbildungen 1 und 2 zeigen die realen Ausbildungspalttformen, deren Entwicklungen in [1] und [2] zu finden sind.

Dieser dritte Teil der Ausbildungsplattformen beschäftigt sich mit der Entwicklung der Steuerungssoftware sowie einer HMI zur Bedienung der Anlage.

Abbildung 3 zeigt die Anforderungen an die Ausbildungsplattform. Dabei werden drei Verwendungsvarianten definiert,

ausgearbeitet und implementiert. Mit der ersten Variante wird es dem Bediener ermöglicht, eine extern, zum Beispiel in *MATLAB*, generierte Trajektorie zu verwenden. Nach dem Export in eine *.csv Datei kann diese über die HMI Applikation geladen und dem System übermittelt werden.

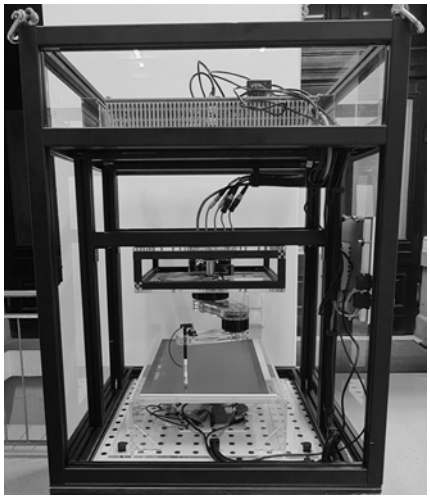


Abbildung 1. 3D-Darstellung der seriellen Kinematik als SCARA Roboter.

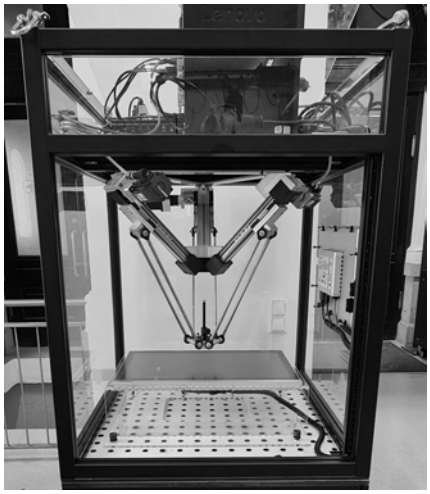


Abbildung 2. 3D-Darstellung der parallelen Kinematik als Delta360 Roboter.

Die zweite Variante ermöglicht die Steuerung der Roboterzellen direkt über *Simulink*. Dabei wird die TE1410 Schnittstelle zwischen *Simulink* und *Beckhoff* verwendet [3]. Das *Simulink* Modell kommuniziert über diese Schnittstelle mit der HMI und ist direkt mit den Variablen der Motorsteuerung verbunden.

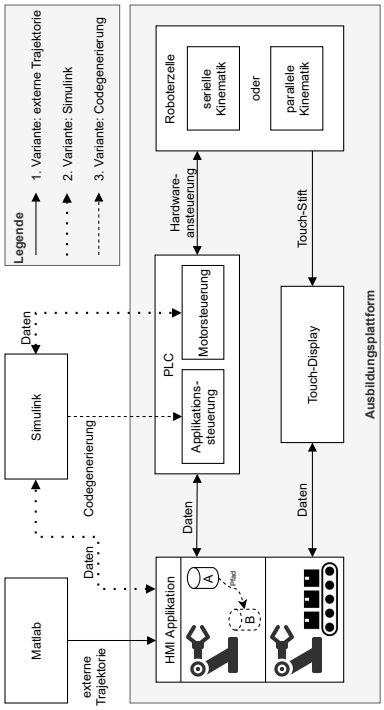


Abbildung 3. Anforderungsanalyse der Roboterapplikation.

Bei der dritten Variante wird das Steuerungsmodell von *Simulink* über die TE1400 Codegenerierung in ein *TcCOM* Module exportiert [4]. Dieses kann anschließend direkt in der Steuerung integriert werden.

III. Konzeptphase

Das Blockschaltbild in Abbildung 4 zeigt den Systementwurf für die Ausbildungsplattform. Dieser Systementwurf teilt sich in drei Hauptkomponenten auf:

- HMI Applikation
- *Simulink* Steuerungsmodell
- PLC

Dabei ist das Laden von externen Trajektorien über die Eingabe des Bedieners in der HMI ersichtlich sowie die Anbindung eines Touch-Displays. Die HMI besitzt zudem noch Anwendungen zur Kalibrierung des Roboters mit dem Touch-Display und zur Ausführung einer virtuellen PP Aufgabe. Zusätzlich beinhaltet die HMI eine TCP/IP Schnittstelle zur Kommunikation mit dem *Simulink* Modell und eine Automation Device Specification (ADS) Schnittstelle zur Kommunikation mit der PLC.

Das *Simulink* Steuerungsmodell besteht aus einer Applikationssteuerung und einer TCP/IP Schnittstelle. Die Applikationssteuerung ermöglicht die Durchführung der zuvor erklärten Anwendungsvarianten. Die TCP/IP Schnittstelle wird für die Kommunikation zur HMI bei der virtuellen Systemsimulation

benötigt. Die TE1410 ADS Schnittstelle ermöglicht die direkte Ansteuerung der Motoren aus *Simulink*.

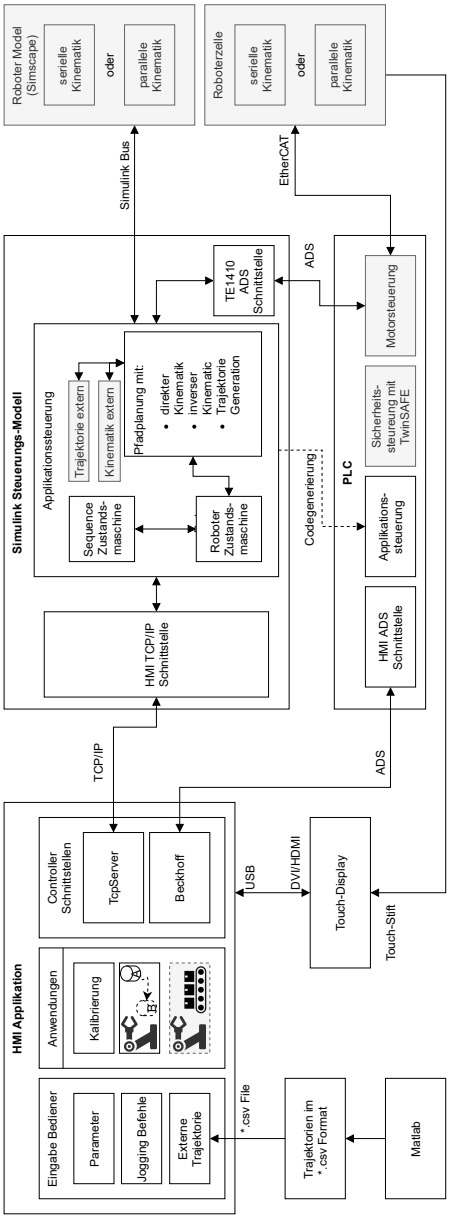


Abbildung 4. Systementwurf der Ausbildungsplattform mit den notwendigen Schnittstellen und Softwarekomponenten.

Für die PLC wird eine *Beckhoff* Steuerung verwendet. Die PLC Implementierung besitzt das *Simulink* Modell als *TcCOM* Objekt und eine ADS Schnittstelle. Bei den grauen Blöcken handelt es sich um die Sicherheitssteuerung und die Motorsteuerung, welche nicht Teil dieser Arbeit sind.

IV. Entwicklungs- und Integrationsphase

Die Ausarbeitung, Entwicklung und Integration des Systementwurfs erfolgt über drei Schritte und dabei wird die VIBN und eine abgeleitete Form der modellbasierten XiL Methode verwendet. Abbildung 5 zeigt die Definition der XiL Methode aus der Richtlinie VDI/VDE 3693 ”Virtuelle Inbetriebnahme - Modellarten und Glossar”.

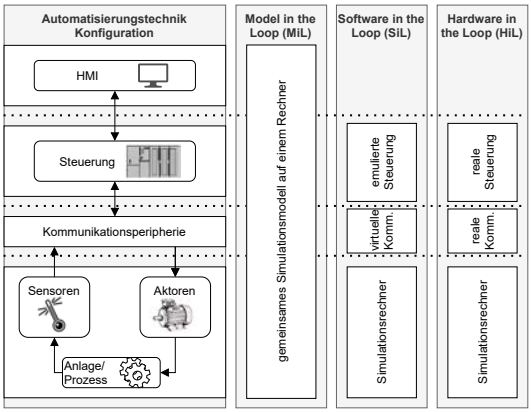


Abbildung 5. Stufen der XiL Entwicklung abgeleitet von der VDI/VDE 3693 [5].

Diese Automatisierungstechnik Konfiguration setzt sich aus der HMI, Steuerung, Kommunikationsperipherie und der Anlage zusammen. Model in the Loop (MiL) behandelt die entwicklungsbegleitenden Tests von Systemmodellen, welche meistens in derselben Simulationsumgebung laufen. Software in the Loop (SiL) dient zum Testen des Seriencodes auf einer emulierten Steuerung mit einem Anlagenmodell. Hardware in the Loop (HiL) beschreibt das Testen des Steuerungscodes auf der Zielhardware, welche ebenso gegen das Anlagenmodell getestet wird [5]. Weitere Anwendung dieser Methode sind in [6], [7] und [8] zu finden.

Abbildung 6 zeigt die drei abgeleiteten Entwicklungsschritte. Der erste Schritt ist die Entwicklung und deckt die MiL Methode ab. Dabei wird die Applikationssteuerung der PLC als Modell in *Simulink* ausgearbeitet. Zusätzlich werden in [1] und [2] virtuelle Robotermodelle in *Simscape* zur Verfügung gestellt, womit die virtuelle Systemsimulation mit *MATLAB/Simulink* ermöglicht wird. Die HMI wird in *C++* mit *Visual Studio* 2019 und der Bibliothek *Qt* programmiert [9].

Der zweite Schritt beinhaltet die VIBN. Dabei wird das Steuerungsmodell mittels TE1400 Codegenerierung in einer *Beckhoff* Steuerung integriert. Anschließend wird der Datenfluss zwischen der HMI, dem *TcCOM* Objekt und der Motoransteuerung in der PLC implementiert. Darauf folgt die Ausarbeitung der ADS Schnittstelle mit *QAds* [10] in der HMI. Das virtuelle Robotermodell kann anschließend über die reale

Steuerung in der PLC gesteuert und über die HMI bedient werden. Dieser Schritt wird mit der SiL Methode validiert.

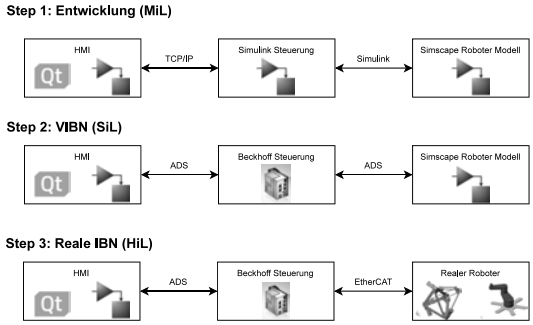


Abbildung 6. Definition der Entwicklungsschritte.

Der dritte Schritt deckt abschließend noch die reale Inbetriebnahme der Ausbildungsplattformen ab und verwendet die HiL Methode. Damit die Ergebnisse der Entwicklungsschritte gegenüber gestellt werden können, wird in *Simulink* eine Testapplikation entwickelt. Diese ermöglicht die Ausführung der selben Abläufe in allen XiL Testdurchläufen.

A. Entwicklung (MiL)

Das Systemmodell in *Simulink* ist in Abbildung 7 zu finden. Dieses Modell der MiL Entwicklung besitzt

- ein HMI TcpInterface,
- eine SafetyHardware Simulation,
- einen Delta/SCARA Controller und
- ein Delta/SCARA Robot Model

und ermöglicht eine virtuelle Systemsimulation. Das Herzstück des Modells ist der Controller. Dieser beinhaltet drei Zustandsmaschinen und die Pfadplanung für die Roboterbewegung.

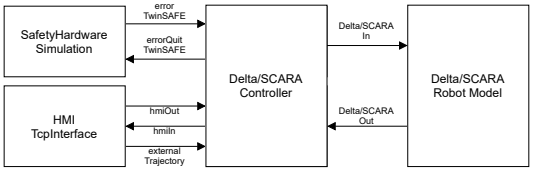


Abbildung 7. Systemmodell in *Simulink* mit HMI Schnittstelle, Sicherheitsmodell, Applikationssteuerung und Robotermodell.

Abbildung 8 stellt die Zustandsmaschine der Applikationssteuerung dar. Diese nimmt Befehle von der HMI entgegen und leitet Bewegungsbefehle an die `RobotControl` Zustandsmaschine weiter. Bei allen Zustandsmaschinen handelt es sich um Mealy Automaten [**?**], die mit der *Stateflow* Toolbox von *MATLAB/Simulink* ausgearbeitet werden [11]. Der Einstiegspunkt befindet sich beim Zustand `Stopped`. Dieser Zustand kann von jedem anderen Zustand über den Zustandswechsel `Stop` erreicht werden. Anschließend wird über `Home` der Roboter über die Zustände `HomingRobot` und `WaitForRobot` auf eine definierte Heimposition bewegt und der Automat landet beim Zustand `Idle`.

Über die `Start...`-Zustandswechsel können die Anwendungen

- Jogging,
- Calibration,
- Automatic und
- External Trajectory

gestartet werden. Bei einem sicherheitskritischen Fehler, welcher beim realen System durch die Sicherheitssteuerung und beim virtuellen durch das SafetyHardware-Modell hervorgerufen wird, erzwingt der Zustandswechsel `safetyError` den Zustand `SafetyError`. Nur wenn kein Fehler mehr vorhanden ist und der Zustandswechsel `errorQuit` durchgeführt wird, kann dieser Zustand wieder verlassen werden. Mehr zur Sicherheitsvorrichtung ist in [1] und [2] zu finden.

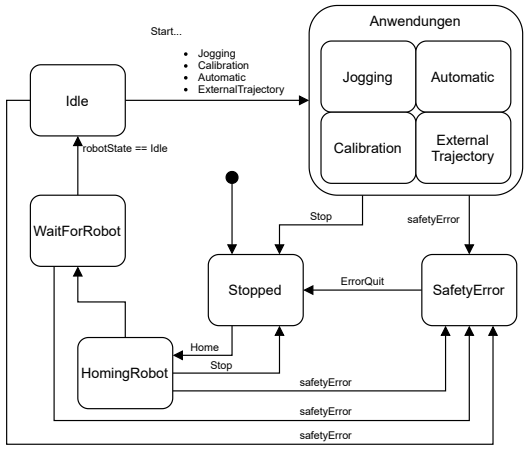


Abbildung 8. *Stateflow* Modell der `SequenceControl` Zustandsmaschine.

Die `RobotControl` Zustandsmaschine, dargestellt in Abbildung 9, nimmt Bewegungsbefehle von der `SequenceControl` entgegen und gibt den aktuellen Befehl an die Pfadplanung weiter. Der Eintrittspunkt liegt hier wie zuvor beim Zustand `Stopped`, welcher über den Zustandswechsel `Stop` von jedem anderen Zustand erreicht werden kann. Nach dem Heimfahren befindet sich der Automat bei `Idle`. Über den `Start` Zustandswechsel landet der Automat bei `Busy`. Dort werden alle vorhandenen Bewegungsbefehle abgearbeitet und anschließend erfolgt der Wechsel zurück auf `Idle`. Sollte der Roboter während der Ansteuerung einen Fehler melden, erfolgt der Wechsel auf `Error`. Mit `Stop` und `Home` wird der Fehler zurückgesetzt und der Roboter ist wieder einsatzbereit.

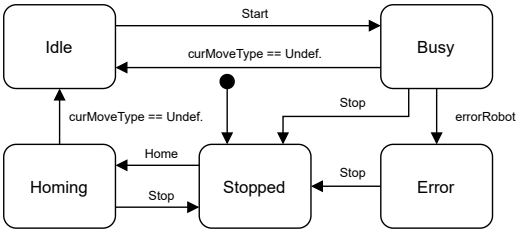


Abbildung 9. *Stateflow* Modell der `RobotControl` Zustandsmaschine.

Neben den Zustandsmaschinen besteht der zweite Teil des Robotercontrollers aus der Pfadplanung. Abbildung 10 zeigt ein vereinfachtes Blockschaltbild des Aufbaus in *Simulink*. Die Eingangsdaten werden von den Zustandsmaschinen bereitgestellt und im `Inputhandler` weiterverarbeitet. Zusätzlich zu den Eingangsdaten werden die aktuellen Roboterdaten sowie die Ergebnisse der direkten Kinematik an den `Inputhandler` weitergereicht. Der `Inputhandler` stellt die benötigten Startwerte für die Berechnung der Trajektorie zur Verfügung.

Je nach Bewegungsbefehl wird in der Trajektorienberechnung mit `MoveL`, `MoveJ` oder `MoveExtern` der aktuell notwendige Wegpunkt mit Geschwindigkeit und Beschleunigung ermittelt und dem `Outputhandler` weitergeleitet. Dieser bearbeitet die Daten nochmals und stellt dem Roboter schlussendlich die Soll-Daten zur Positionierung der Achsen zur Verfügung. Die Ausarbeitungen der einzelnen Blöcke sind vom Robotertyp abhängig, weshalb auf die detaillierte Implementierung in [1] und [2] verwiesen wird.

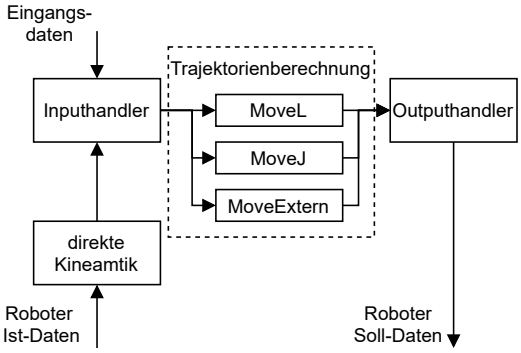


Abbildung 10. Vereinfachtes Blockschaltbild der Pfadplanung.

Nach der Entwicklung des Steuerungsmodels folgt die Ausarbeitung der HMI, welche in Abbildung 11 ersichtlich ist. Die Hauptansicht besitzt eine Toolbar mit verschiedenen Buttons, womit die einzelnen Zustandswechsel der `SequenceControl` getriggert werden können. Beginnend von rechts handelt es sich um `Home`, `Stop`, `StartJogging`, `StartCalibration`, `StartAutomatic`, `StartExternalTrajectory` und `ErrorQuit`. Die Combobox ermöglicht die Auswahl verschiedener Automatikaufgaben.

Abbildung 11. Anzeige des Hauptfensters der HMI.

Im Hauptfenster selbst sind die Bedienelement für Jogging, Parametereinstellung und externe Trajektorie ersichtlich. Je nach Zustand der `SequenceControl` werden diese aktiviert und deaktiviert. In der Statusbar befinden sich aktuelle Informationen der verbundenen Ausbildungsplattform.

Die Parameter dienen zur Begrenzung der Geschwindigkeiten und Beschleunigungen sowie der maximalen z-Position. Ebenso wird die Heimposition des Roboters definiert. Mit Jogging kann der Endefektor des Roboters linear über eine auswählbare Distanz im Raum bewegt werden. Das ermöglicht eine manuelle Steuerung des Roboters. Die Kalibrierung berechnet den Abstand zwischen der Heimposition des Roboters und dem linken oberen Bildrand des internen Touch-Displays. Ebenso wird die Ausrichtung ermittelt, wodurch eine Positionierung des Endefektors am Display möglich ist.

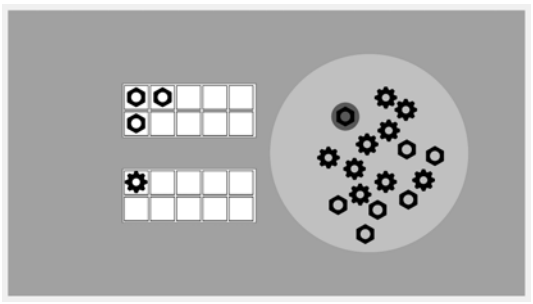


Abbildung 12. Darstellung der Automatikanwendung Gears and Nuts to Tray.

Abbildung 12 zeigt eine mögliche PP Aufgabe des Automatikmodus. Dabei werden die Muttern und Zahnrädern im hellgrauen Kreis den beiden Rastern auf der linken Seite zugeordnet. Durch die Touch-Eingabe des Endefektors am Display kann ein Objekt selektiert werden. Durch ein weiteres

Touch-Event wird ein selektiertes Objekt auf die neue Position verschoben, wodurch das Aufnehmen und Ablegen simuliert wird.

Die Verwendung der externen Trajektorie greift ebenso auf die Automatikanwendung zurück. Dabei werden dem Bediener die aktuellen Achspositionen sowie die PP Positionen zur Verfügung gestellt. Dieser kann anschließend anhand der Daten die Trajektorie berechnen, laden und starten. Stimmt die externe Trajektorie mit den gegebenen Positionen überein, wird das Objekt wie im Automatikablauf aufgenommen und im Raster abgelegt. Diese externe Trajektorie ist mit 1000 Wegpunkten limitiert und die Spalten in der notwendigen *.csv Datei besitzen die Variablen t, $q1$, $q2$ und $q3$ in genau dieser Reihenfolge.

Für die Kommunikation mit dem Steuerungsmodell besitzt die HMI einen TCP/IP Server. Beim Starten des *Simulink* Modells verbindet sich dieses als Client mit diesem Server und ermöglicht den Datentransfer.

B. VIBN (SiL)

Bei der VIBN wird das Modell in eine ausführbare Software umgewandelt und die Funktionalität auf der PLC getestet. Dafür wird die TE1400 Codegenerierung verwendet, wodurch aus dem Modell ein *TcCOM* Objekt erzeugt wird. Diese wird in die Steuerung integriert und mit einer definierten Zykluszeit ausgeführt.

Anschließend wird der Datenfluss, zu sehen in Abbildung 13, ausgearbeitet und implementiert. In der Mitte befinden sich die globalen Variablenstrukturen `Control` und `Maestro`. Über die `Control` Struktur werden die Daten zwischen *Simulink*, HMI Applikation und dem `PRG_Controller` ausgetauscht. Die `Maestro` Struktur wird für den Austausch der Achsdaten zwischen *Simulink*, `PRG_Controller` und `PRG_Maestro` verwendet.

Der `MAIN` Task kümmert sich um den Aufruf der drei Programme. Dabei beinhaltet das `PRG_GenAds` die notwendige ADS Schnittstelle zu *Simulink*. In *Simulink* selbst wird das TE1410 Interface für diese ADS Kommunikation benötigt [3]. Der Datenaustausch erfolgt dabei über Array-Variablen.

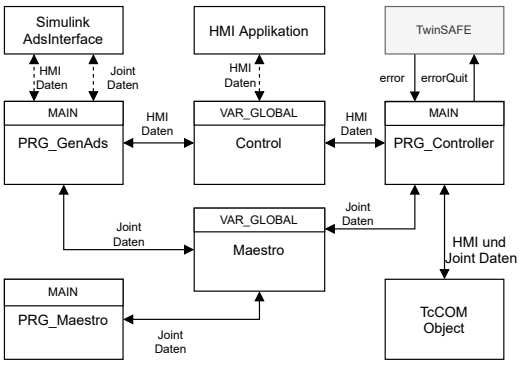


Abbildung 13. Blockdarstellung des Steuerungsaufbaus inklusive der Datenflussmöglichkeiten.

Das `PRG_Controller` ist für die richtige Zuweisung der HMI und Joint Daten, je nach Verwendung, verantwortlich. Dabei wird unterschieden, ob die Ansteuerung der Achsen über *Simulink* oder das *TcCOM* Objekt durchgeführt wird.

Abschließend beinhaltet der `MAIN` Task das `PRG_Maestro`. Dieses ist für die Ansteuerung der Achsen verantwortlich und tauscht dementsprechend mit der `Maestro` Struktur die Achsdaten aus.

Diese Implementierung des Datenflusses ermöglicht verschieden Anwendungsmöglichkeiten der Ausbildungsplattform. Fürs Erste kann über die HMI Applikation das *TcCOM* Objekt und somit der Roboter gesteuert werden. Des Weiteren wird mit der *Simulink* Schnittstelle die Steuerung des Roboters inklusive HMI über das Steuerungsmodell in *Simulink* ermöglicht.

Für die SiL Überprüfung wird mit dem *TcCOM* Objekt über eine HMI Testapplikation in *Simulink* das Robotermodell in *Simscape* gesteuert. Diese HMI Testapplikation wird ebenso im vorherigen Schritt verwendet, wodurch die Ergebnissen gegenübergestellt werden können.

C. Reale Inbetriebnahme (HiL)

Als Ergebnis der VIBN und der SiL Überprüfung steht ein getestetes Software Framework mit HMI zur Verfügung, welches das virtuelle Robotermodell steuert. Für die reale Inbetriebnahme fehlt noch die Ansteuerung der Achsen, welche im `PRG_Maestro` implementiert ist.

Dabei werden die Achsen über die NC Motion Implementierung in der PLC mit einer Geschwindigkeitsregelung integriert. Die Positionsregelung zur Trajektorie mit gegebener Geschwindigkeit wird mit einem Feedforward Regler in der Steuerung selbst hinzugefügt. Mehr zu diesen Entwicklungen ist in [1] und [2] zu finden.

Abschließend wird die reale Hardware in der HiL Überprüfung mit der HMI Testapplikation über *Simulink* gesteuert, wodurch die Ergebnisse mit MiL und SiL gegenübergestellt werden können.

V. Anwendungsfälle zum Test des Frameworks

Mit den Anwendungsfällen des Delta360 und SCARA Roboters werden die entwickelten Anwendungsvarianten getestet. Für die Bedienung der Ausbildungsplattform in *MATLAB/Simulink* steht ein `RunAusbildungsplattform` Skript zur Verfügung. Dieses ermöglicht

- die virtuelle Systemsimulation,
- die XiL Überprüfung,
- die Steuerung des Roboters über *Simulink*,
- die Codegenerierung und
- die Durchführung einer Sprungantwort.

Bei der virtuellen Systemsimulation wird abhängig vom Robotertyp das zugehörige *Simulink* Modell und die HMI gestartet. Anschließend kann die virtuelle Anlage bedient und verwendet werden. Abbildungen 14 und 15 zeigen diese Systemsimulation während der Automatikanwendung.

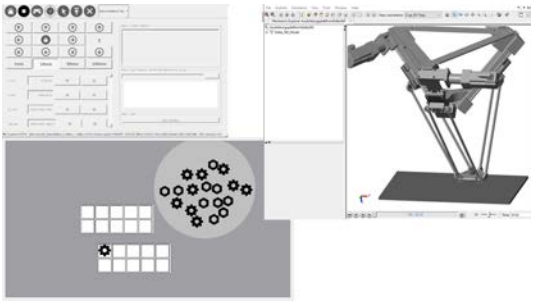


Abbildung 14. Virtuellen Systemsimulation des Delta Roboters.

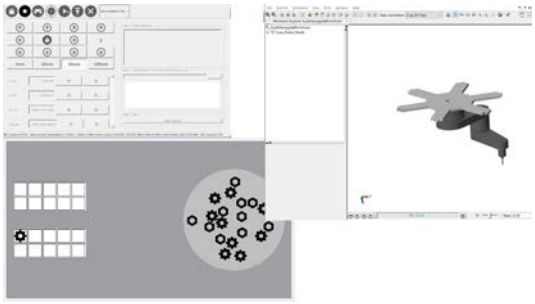


Abbildung 15. Virtuellen Systemsimulation des SCARA Roboters.

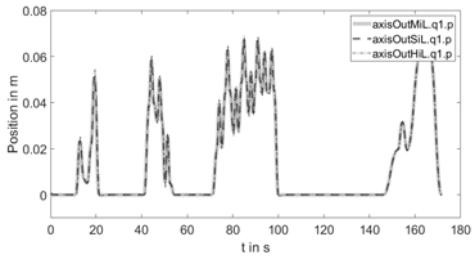


Abbildung 16. Vergleich der Positionen der XiL Testdurchläufe des Delta Roboters.

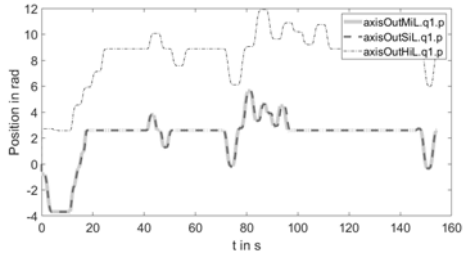


Abbildung 17. Vergleich der Positionen der XiL Testdurchläufe des Delta Roboters.

Abbildung 18. Positionierung im Kalibrier- und Automatikablauf des realen Delta Roboters.

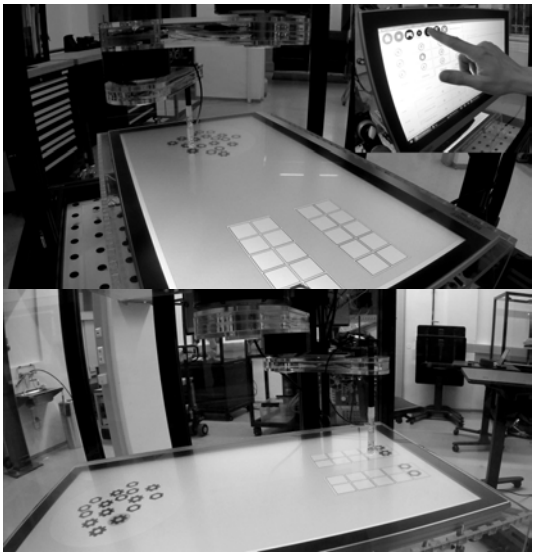


Abbildung 19. Positionierung im Automatikablauf des realen SCARA Roboters.

Die Ergebnisse der XiL Überprüfung in den Abbildungen 16 und 17 bestätigen dieselbe Funktionalität des Modells und der Steuerung. Die Delta360 Ergebnisse weichen nur gering voneinander ab. Beim SCARA ist der Offset des HiL Resultats von $2\pi\,\mathrm{rad}$ sowie die zeitliche Verzerrung erkennbar. Diese ist auf die Initialisierung der Achsen und der Positionsregelung der Motoren zurückzuführen.

Die Steuerung des realen Roboters über *Simulink* erfolgt über die TE1410 Schnittstelle und ermöglicht dieselben An-

wendungen wie jene vom Systemmodell beim virtuellen Roboter. Diese Funktion ermöglicht das Testen von Entwicklungen am realen System ohne Codegenerierung.

Über die TE1400 Codegenerierung wird die Verwendung der realen Ausbildungsplattform ohne *MATLAB/Simulink* ermöglicht. Dabei wird die Steuerung der *Beckhoff* PLC über *TwinCAT* aktiviert und gestartet. Durch das Starten der HMI über den Shortcut wird diese richtig mit der Steuerung verbunden, da über den Shortcut die Route zur PLC definiert ist. Die HMI aktiviert beim Starten das `PRG_Controller` und `PRG_Meastro`. Darauf folgt die Initialisierung der Achsen und die Anlage ist betriebsbereit. Abbildungen 18 und 19 zeigen Ausschnitte der Positionierungen der Roboter.

VI. Zusammenfassung und Ausblick

Diese Arbeit zeigt die schrittweise und modellbasierte Entwicklung des Software Frameworks für die Steuerung der Ausbildungsplattform mit serieller und paralleler Kinematik. Dieses Software Framework besteht aus einer *Simulink* Bibliothek mit Systemmodellen, einer Steuerungsapplikation in *TwinCAT* für eine *Beckhoff* PLC und einer HMI entwickelt in *C++* mit *Qt*.

Die Entwicklung des Systemmodell, über die Integration in die PLC bis hin zur Ansteuerung des realen Robotersystems ist getestet und die Funktionen sind verifiziert. Das resultierende Anlagenmodell ermöglicht die virtuelle Systemsimulation, wodurch erarbeitete Lösungen von Kinematik und Trajektorienplanung ortsunabhängig getestet werden können. Ebenso ist die Bedienung dieser Systemsimulation mit der resultierenden HMI möglich.

Das Steuerungsmodell wird über die TE1400 Codegenerierung in *TwinCAT* integriert. Dadurch besitzt das reale System dieselben Anwendungsmöglichkeiten wie das virtuelle. Nach erfolgreicher VIBN der erarbeiteten Lösungen können diese mit der Ausbildungshardware getestet werden.

Grundsätzlich ist das Software Framework ohne zusätzlichen Aufwand vollständig funktionsfähig. Die HMI ermöglicht die Bedienung der Ausbildungsplattformen und bietet eine manuelle Jogging-Funktion, eine Automatikanwendung mit Kalibrierung und das Laden externer Trajektorien an.

Prinzipiell eignet sich die Entwicklung dieser Ausbildungsplattform als Demonstrationsbeispiel der XiL Methode. Voraussetzungen sind gute Kenntnisse in der Verwendung von *MATLAB/Simulink* und *TwinCAT*. Sind diese Kenntnisse gegeben, kann die Ausbildungsplattform ebenso für die Entwicklung einer Applikationssteuerung für Automatisierungsanlagen inklusive Roboterkinematik verwendet werden.

Eine Schwäche des Systems liegt in der Anwendung von *Beckhoff* selbst. Die Steuerung muss bei der Verwendung der Anlage über *TwinCAT* aktiviert und gestartet werden. Wenn Ingenieure diese Schritte in *TwinCAT* nicht selber erledigen sollen, müsste diese über die Automation Interface, zu finden in [12], automatisiert durchgeführt werden.

Der nächste Schritt besteht in der Anwendung der Ausbildungsplattform. Dabei wird Feedback zur Bedienung und Anwendbarkeit gesammelt und auf Basis dessen Verbesserungen abgeleitet.

Mögliche Erweiterungen zum System wären die Integration eines Bildverarbeitungssystems sowie die Erweiterung der PP Anwendungen um eine Förderbandsimulation. Ebenso das Zeichnen von Vektorgrafiken mittels Laser und fluoreszierender Folie wäre eine zusätzliche Funktionserweiterung.

Danksagung

Der Autor möchte sich beim MCI für das Thema der Arbeit und bei seinen Kommilitonen Serafin Kolleger und Christoff Sulzenbacher für die aktive Zusammenarbeit während der Entwicklung dieser Ausbildungsplattform bedanken.

Literatur

[1] C. Sulzenbacher, “Entwicklung einer Ausbildungsplattform für serielle Kinematiken,” *Management Center Innsbruck*, 2021.

[2] S. Kollegger, “Entwicklung einer Ausbildungsplattform für parallele Kinematiken,” *Management Center Innsbruck*, 2021.

[3] Beckhoff, “Handbuch - TC3 Inteface für Matlab/Simulink,” https://download.beckhoff.com/download/document/automation/twincat3/TE1410_TC3_Interface_Matlab_DE.pdf, 2017, [Online; besucht am 21.06.2021].

[4] ——, “TC3 Target Inteface für Matlab/Simulink,” https://download.beckhoff.com/download/document/automation/twincat3/TE1400_TC3_Target_Matlab_DE.pdf, 2020, [Online; besucht am 21.06.2021].

[5] VDI/VDE-Fachbereich Engineering und Betrieb, *VDI/VDE 3693 Blatt 1: Virtuelle Inbetriebnahme - Modellarten und Glossar*. VDI/VDE-Gesellschaft Mess- und Automatisierungstechnik, 2016.

[6] Dr.-Ing. Christian Henke, “X-in-the-Loop Entwicklungs- und Testumgebungen,” https://www.iem.fraunhofer.de/content/dam/iem/dokumente/leistungsangebote/Regelungstechnik/leistungsangebot-xil-entwickluns-und-testumgebungen-iem.pdf, 2021, [Online; besucht am 02.03.2021].

[7] MathWorks Support Team, “What is MIL, SIL, PIL, HIL and how do they integrate in Model Based Design approach?” https://de.mathworks.com/matlabcentral/answers/440277-what-is-mil-sil-pil-hil-and-how-do-they-integrate-in-model-based-design-approach, 2021, [Online; besucht am 15.04.2021].

[8] Beckhoff, “TE1400 TwinCAT Target für MATLAB/Simulink,” https://infosys.beckhoff.de/index.php?content=../content/1031/te1400_tc3_target_matlab/9007203465956235.html&id=, 2021, [Online; besucht am 15.04.2021].

[9] Qt, “One framework. One codebase. Any platform - About Us,” https://www.qt.io/company, 2021, [Online; besucht am 29.04.2021].

[10] Stefan Besler, “QAds,” https://github.com/stefanbesler/QAds, 2021, [Online; besucht am 09.06.2021].

[11] MathWorks, “Stateflow - Model and simulate decision logic using state machines and flow charts,” https://www.mathworks.com/products/stateflow.html, 2021, [Online; besucht am 10.06.2021].

[12] Beckhoff, “Handbuch - Automation Interface,” https://download.beckhoff.com/download/document/automation/twincat3/Automation_Interface_DE.pdf, 2020, [Online; besucht am 06.09.2021].

Julian Muigg ist Student am *Management Center Innsbruck (MCI)* im Master für Mechatronik & Smart Technologies. Er arbeitet hauptberuflich in der Optronik-Abteilung bei *Swarovski Optik* in Absam.

Design, Simulation and Analysis of a Permanent Magnet Synchrounous Motor

Lukas Nuschele and Davide Bagnara (supervisor)

***Abstract*—Due to the ongoing climate change and rising CO_2 emissions, sustainable drive systems for all kinds of vehicles have become increasingly important. Within the whole construction machine sector, a high demand for more sustainable paths and investments arises.**
In this specific master thesis an electric motor of a crawler tractor was designed, simulated and analyzed using the software Ansys. The interior permanent magnet synchronous motor (PMSM) was designed with similar torque/speed characteristics as the existing crawler tractor, powered by a diesel engine. Therefore, it was at first important to analyze the alternative drive system, in order to define the extended operating range. After that, the motor was designed on an analytical basis. This enabled to create a first geometrical design with the free online software *JMAG*. With the analysis results from *JMAG*, more accurate designs could be generated in an iterative process, using *Ansys Electronics*.
After the motor was designed, an electromagnetic simulation, on the basis of finite element analysis, was carried out. This allowed to analyze all important characteristics of the motor. In order to optimize the performance of the motor, many different design features and excitation strategies were simulated. On top of that, the thermal properties of the motor were investigated by a coupled simulation with *Ansys Mechanical*.
In the end, the outcome together with future related possibilities in electric motor design is discussed and concluded.

***Index Terms*—PMSM, Motor design, Electromagnetic simulation, Thermal analysis, Ansys Electronics, FEM**

I. Introduction

ACCORDING to MAHB (Millennium Alliance for Humanity and the Biosphere), initiaded by the Stanford University, the worldwide oil resources will end up in the year 2052 [1]. Obviously the development of electrified means of transport is inevitable. The faster this electrification process is happening, the better it is for the global climate and environment.
Therefore an efficient design for alternative drive systems of heavy duty vehicles has to be developed. This comprises the research in productive battery systems and green energy provision, for example with fuel cells. One very important part in the electric drive development is the motor design. An accurate and appropriate motor design is crucial to save costs, reduce environmental pollution and maximize its characteristics. In this thesis, a permanent magnet synchronous machine (PMSM), as part of an electrification process for an electric crawler tractor, was designed, simulated and analyzed.
The design of a proper model of a motor is very complex and can't be fulfilled by following straight forward standards, as there are too many parameters effecting the outcome. In order to design a motor model with satisfying characteristics, the whole design phase is an iterative process. Lots of simulations were carried out in order to analyze different settings and according outcomes. Therefore it is very important to keep different factors in mind. Especially economic factors , such as manufacturing costs, energy consumption, choice of material and performance, influence the design process. Also material limitations always have to be considered [2].
The starting point of the motor creation was based on the torque/speed characteristics of the diesel driven engine. Following curve 1 determines these characteristics.

Lukas Nuschele studies at the Management Center Innsbruck, e-mail: lukas.nuschele@gmx.de

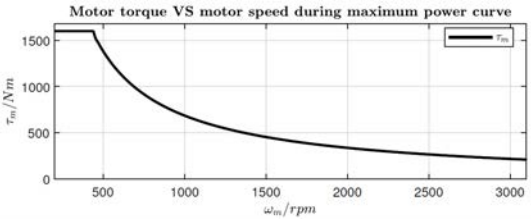


Fig. 1. Performance characteristics of the hydrostatic power train output for one drive chain [3]

A maximum torque of 1.6 kN up to approximately 480 rpm had to be achieved for one side of the drive chain. The operating range was defined to 3100 rpm at a minimum torque of 230 kN. These parameters enabled to define first geometrical parameters, such as the the outer diameter of the machine.

II. Analytical Motor Design

Due to space and weight limitations the motor shall be kept as small and as light weighted, as possible. As a constraint, a maximum outer diameter D_{st} of 600 mm with a maximum height of 160 mm is considered. The outer diameter of the stator is strongly affected by the number of poles in the motor. When using higher number of poles, the stator and rotor yoke thickness can be decreased [4]. A reduction of the outer diameter goes along with the advantages of lower space requirement and less motor weight. In image 2 the relation of poles to the outer stator diameter is depicted.

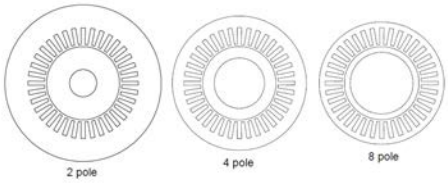


Fig. 2. Effect of different number of poles to the outside diameter of the stator [4]

The lower the pole numbers, the higher is the outside diameter of the motor. Because of the space limitations in the heavy duty vehicle, the 8 pole configuration is considered as a starting point.
Further important design features of the interior permanent magnet synchronous machine were based on analytical calculations. Since the maximum output voltage of the battery is constrained to $\hat{V}_{DC} = 800$ Volts, the maximum phase peak voltage can be calculated as follows:

$$\hat{E_{ph}} = \frac{\hat{V}_{DC}}{\sqrt{3}} = 461.88V \tag{1}$$

Since the number of poles is set to 8 and the maximum mechanical speed is also known, the magnetic flux value is

$$\hat{\psi}_M = \frac{\hat{E_{ph}}}{p\hat{\omega_M}} = 0.356Wb. \tag{2}$$

This quantity is very important to determine the magnet sizes and also stands in relation to the maximum occurring phase currents. This property is calculated with following equation.

$$\hat{i}_{phase} = \frac{2\hat{\tau}_m}{3p\hat{\psi}_M} = 749,06A \tag{3}$$

In order to define further geometrical sizing parameters, next step was to select the motors materials. The material of choice for the stator and rotor was set to laminated electrical steel, M800-65A. The biggest advantage of this material is a high heat resistance, due to its laminated structure. Additionally M800-65A convinces by a high permeability and low core losses [5]. The substance for the magnet was set to N42-H (NdFeB, Neodynium Iron Boron), which is one of the strongest materials for permanent magnets worldwide. Very high remanence and coercivity distinguishes it as the perfect suited magnetic material, creating a strong magnetic flux [6]. The winding of the electric machine can either be developed out of copper or aluminum wire. Copper is one of the best suited and common materials for the creation of motor windings, since it has a high electric conductivity and good thermal heat conductivity properties. However, aluminum is much cheaper in terms of material prices and also lighter compared to copper [7]. In this thesis both configurations were analyzed and compared.
Furthermore, the number of slots has to be defined. According to an investigation on different slot designs for a motor with similar power characteristics in [8], a configuration with 48 slots was chosen.

With the knowledge of the material properties and further assumptions the rest of the most important geometrical quantities could be calculated. One important calculation was the determination of the air gap flux density B_g, which is

$$B_g = \frac{\psi_m 4p}{D_r \pi L_{stack} N_{turns}} = 0.9373T \tag{4}$$

The parameter D_r stands for the rotor diameter of 403 mm and N_{turns}, for the number of winding turns. This value was set to 30 according to a study on the influence of different number of winding turns in a PMSM [9]. The calculation of the air gap flux density enabled to determine the magnet size. The length of the mangent can be calculated as

$$h_m = \frac{B_g \pi D_r}{4pB_m} = 78.1mm \tag{5}$$

whereas the width is defined as

$$t_m = \frac{H_g 2g}{H_m} = 11.9mm. \tag{6}$$

The operating quantities B_m and H_m are defined due to the magnet materials hysteresis curve. The parameter H_m stands for the operating magnetic field strength, whereas B_m indicates the operating flux density. The variable g is the air gap distance and was assumed to be 2 mm. The property H_g is the magnetic field strength in the gap and can be interpolated out of the magnetic flux density in the gap. All these calculations were necessary to generate a first digital pre-design.

III. Predesign with JMAG

For a first pre-design of the motor, the free online software *JMAG* was used. This software is provided by the japanese JSOL corporation and offers parameter based motor design. The target values and initial sizing parameters for the simulation in *JMAG* were calculated analytically and are listed in the following table I.

TABLE I
TARGET VALUES AND INITIAL SIZING PARAMETERS IN *JMAG*

	Parameter	Definition	Value	Unit
Target	P	Rated Power	80	kW
	τ	Max. torque	1600	Nm
	ω_{mech}	Max. rev. speed	3100	rpm
Sizing	p	Number of poles	8	
	N_{slots}	Number of slots	48	
	i_{RMS}	Max. current (RMS)	530	A
	D_{st}	Outer stator diameter	600	mm
	L_{stack}	Max. motor height	160	mm

In the next table II the most important dimensional settings are shown.

TABLE II
DIMENSIONAL SETTINGS FOR THE *JMAG* DESIGN

	Parameter	Definition	Value	Unit
Stator	g	Air gap length	2	mm
	$D_{innerSt}$	Inner Stator Diameter	403.1	mm
	d_{th}	Tooth width	12.7	mm
	B_{s0}	Slot opening width	3.879	mm
	d_{cbw}	Core back width	30.14	mm
	H_{s0}	Tooth tangular depth	4.018	mm
Rotor	D_r	Outer Rotor Diameter	399.1	mm
	D_{shaft}	Shaft Diameter	240	mm
	t_m	Thickness of the magnets	11.9	mm
	h_m	Length of the magnets	40	mm
	ϕ	Angle of the magnets	120	deg

Following table III includes the settings for the winding configuration.

TABLE III
SETTINGS FOR THE WINDING CONFIGURATION

	Parameter	Definition	Value	Unit
Winding	N_{par}	Number of parallels	4	
	N_{turns}	Number of turns	30	
	η	Slot fill factor	70	%
	R_{ph}	Phase resistance	0.0118	Ω
	N_{pitch}	Winding pitch	5	

With these settings the initial *JMAG* design could be generated. This is shown in image 3.

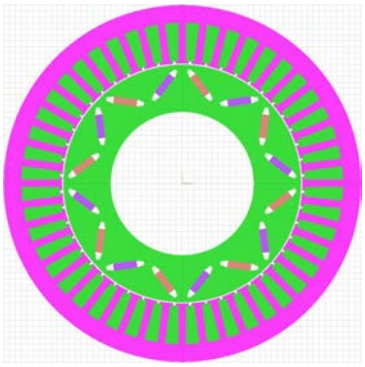


Fig. 3. Initial PMSM design in *JMAG*

This simulation gave important insights in the design structure, which had to be created afterwards in *Ansys Maxwell*. Furthermore, it approved that the motor reaches the necessary torque-speed characteristics and information about the occurring voltage ranges could be plotted. In order to accelerate the design process, this simulation was only conducted with a copper winding configuration. With the initial *JMAG* design, it was possible to generate a PMSM design with the *Ansys* embedded software tool called *Rmxprt*.

IV. RMXPRT DESIGN

Beginning with the machine type selection, necessary parameters for the stator, rotor, windings, core and slots are selected. In table IV, the rest of the main sizing parameters are listed.

TABLE IV
GEOMETRICAL VALUES FOR POSITIONING THE MAGNET AND SLOT DESIGN

	Parameter	Definition	Value	Unit
Magnet Position	d_{mag}	Magnet position	150	mm
	O_1	Inner clearance	18	mm
	Rib	Outer clearance	15	mm
	$Hrib$	Flux barrier width	12.7	mm
	Φ	Angle of the magnets	129.6	deg
	d_{cbw}	Core back width	30.14	mm
Slot	H_{s0}	Tooth tangular depth	4.01	mm
	B_{s0}	Slot opening width	3.9	mm
	B_{s1}	Inner slot distance	17	mm
	B_{s2}	Outer slot distance	21	mm

Together with the parameters from table I, II and III, the *Rmxprt* design could be generated as shown in image 4.

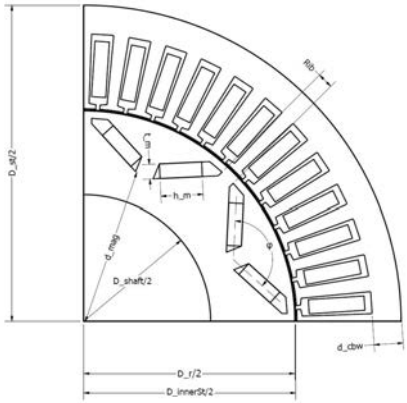


Fig. 4. CAD drawing of a quarter of the PMSM

Next image 5 allows to get a more detailed view of the slot geometry and geometrical position of the magnet flux barriers.

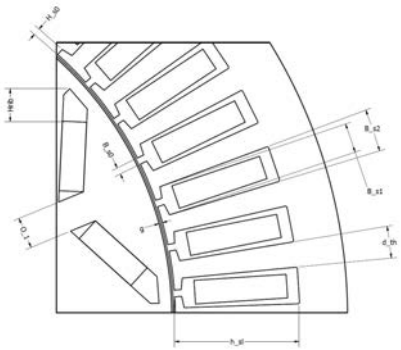


Fig. 5. Detailed view of slot geometry and magnet position

After the *Rmxprt* geometry was designed successfully, main simulation parameters have to be set in order to generate

a *Maxwell* design for a detailed electromagnetic simulation. Following settings depicted in table V were chosen for the first simulation.

TABLE V
Rmxprt SIMULATION SETTINGS

Definition	Value	Unit
Load Type	Constant Speed	
Rated Output Power	80	kW
Rated Voltage (RMS)	179	V
Rated mechanical Speed	500	rpm
Operating Temperature	75	$^{\circ}C$

The RMS voltage value is defined, using the results from *JMAG*. The operating temperature is set, according to the demagnetization curves of the N42H magnets. The maximum temperature is 120 $^{\circ}C$ but at an operating temperature of 75$^{\circ}C$, proper magnetic characteristics are guaranteed [6].
In order to create a Maxwell design, it is necessary to run an analysis in *Rmxprt* at first. With this analysis the phase inductances, resistances and also the weight of the motor can be determined. Following table VI compares the motor weight of a winding configuration with copper to a configuration with aluminum.

TABLE VI
WEIGHT OF THE SINGLE MOTOR PARTS

Part	Copper Design	Aluminum Design
Stator Winding	121.624 kg	36.611 kg
Stator Core Steel	108.843 kg	108.843 kg
Rotor Core Steel	81.826 kg	81.826 kg
Rotor Magnet	10.628 kg	10.628 kg
Overall weight	**322.921 kg**	**237.908 kg**

The *Rmxprt* design comes close to the final motor design but has to be optimized regarding excitation, winding and geometry. One very important factor in motor design is the selection of the winding material. As already described by table VI, the weight of the motor depends highly on the stator winding. Also the manufacturing price is affected strongly by this selection. According to [10], the price for copper in July 2021 was 3.8 times higher as for aluminum. Since the motor shall be used for the production of a series of heavy duty vehicles, this goes along with immense costs. Way more mass in copper would be needed, compared to aluminum since it is way lighter. On the other hand the drawbacks of an aluminum winding are higher losses, due to higher material resistivity and less thermal conductivity [11]. If an identical design for both winding configurations would be used, excessive power losses in the aluminum winding would result. Therefore, it makes sense to use a slightly different design for the aluminum winded motor in order to reduce the current in the stator. Due to these circumstances, two final designs with copper winding and aluminum winding were configured with *Maxwell*. Both results were satisfying but as the configuration with aluminum is the more cost saving variant, it was chosen as the final design. In next chapter the solution of the electromagnetic analysis is described in detail.

V. ELECTROMAGNETIC SIMULATION WITH MAXWELL

Next part of the thesis was to create a 2D Maxwell design of the motor, using the Rmxprt model. Maxwell is an embedded software of *Ansys Electronics*, used for simulating and solving electromagnetic fields of electrical machines. The solver is based on the concept of finite element analysis (FEM), enabling to analyse all important properties such as the flux, current, torque or power.
At first a magnetostatic analysis was conducted in order to check the distribution and range of the flux density in the motor. Therefore, the PMSM is analysed statically without any excitation. This is performed to determine the distribution of the flux lines and the amount of flux density in certain areas such as the air gap. In image 6 the magnetostatic plot of the whole PMSM is depicted.

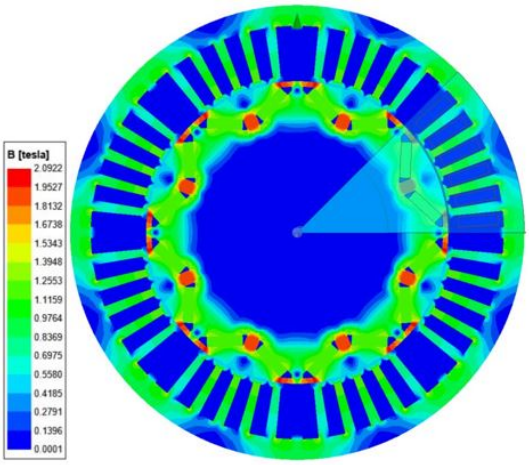


Fig. 6. Distribution of flux density with according scale

The amount of the air gap flux density ranges approximately from 0.5 T up to 0.7 T. The average flux density in the rotor core ranges from 0.8 to 1.3 T. The average flux density in the stator core varies from 0.5 T up to 0.8 T. The magnets itself emit magnetic flux at 1.3 T, indicated by the green color. The highest value of 2.0922 T occurs in between the magnet duct channels, as this is the area where the magnet flux lines are bundled up most. In the iron sector with flux densities above 1.3 T, the iron losses result, since the material is consigned by magnetic saturation. These areas are depicted in red and yellow color.

Most of the iron area is below the value of magnetic saturation. This confirms an equally distributed magnetic flux in the electric machine. The next step in the design phase of the PMSM was the conduction of a transient analysis.
In the transient analysis, the PMSM can be excited with a sinusoidal distributed current, respectively a sinusoidal voltage. This leads to a revolution of the motor and enables a detailed analysis of all important parameters. Therefore, the excitation of the phases has to be set.

At first the simulation at nominal speed is described shortly. The motors phases are excited by a current at nominal speed of 500 rpm. The amplitude of the current signal was evaluated in an iterative process. With an increased magnet size, a maximum current amplitude of i_{max}= 450 A, is necessary to get to the required torque limit. For the phase A, following excitation formula is generated :

$$I_A = i_{max} \sin(2\pi f \cdot t + \gamma) \tag{7}$$

The excitation formula for the phases B and C are the same, only the phase is shifted by +120 degree, respectively -120 degree. The parameter t indicates the simulation time, f the frequency and γ the initial excitation angle of 55°. The initial excitation angle was determined by an analysis with the *Maxwell* embedded software tool, called *Optimetrics*. The frequency of the transient simulation can be calculated as

$$f = \omega_{mech} \frac{p}{2} = \frac{500rpm}{60s} \frac{8}{2} = 33.33Hz. \tag{8}$$

The simulation time was set to 300 ms with a sampling time of $t_s = 0.0001s$. For a better visibility the generated current signal is depicted over 60 ms and shown in following plot 7.

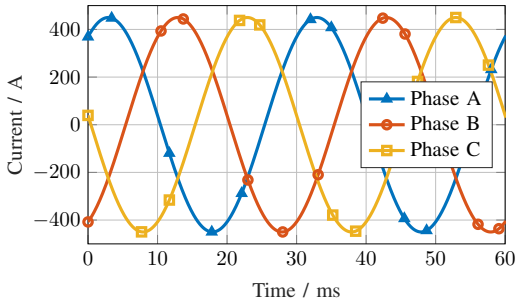


Fig. 7. Input Current signal for 500 rpm, nominal speed simulation

The input current leads to a revolution of the motor at constant speed with a certain magnitude of torque. The output torque signal is depicted in following image 8.

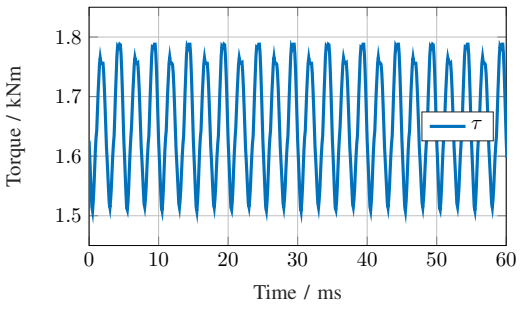


Fig. 8. Output torque τ of the pmsm

The average output torque of the current excitation is at $\tau = 1.658kNm$. It gets clear, that the output torque reaches the desired limit of 1600 Nm at nominal speed. The torque ripple can be approximated to $\pm$ 130 Nm to the average torque. The dominant torque ripple harmonic during simulation is the 12th. This can be noticed in the plot, since 12 full torque ripple cycles are present during one excitation cycle. According to [12], the 12th harmonics, followed by the 6th harmonics , are the most dominant occurring torque ripple harmonics in an electric motor. The resulting frequency of torque ripple acting on the vehicle can be calculated as

$$f_{ripple} = \frac{500}{60} p \cdot 12 = 400Hz. \tag{9}$$

At a ripple frequency of 400 Hz, the heavy duty vehicle won't have problems with machine performance, since it is constructed out of heavy steel parts. Furthermore, these vehicles normally operate on soft ground like earth or gravel, acting as a damping factor during operation.
According to [13], the undesirable effect of torque ripple is caused by deviations of the sinusoidal flux density distribution in the airgap. Furthermore, the selection of the winding geometry and variable magnetic reluctance in the airgap, affect the ripple of the torque signal. The reluctance, or magnetic resistance, in the airgap alternates periodically when the stator teeth pass the edges of the rotor magnets. Reducing the ripple, due to this effect, is possible by skewing the stator slots or by the implementation of a current control structure with additional torque ripple controller.
In order to implement a proper current control system, following data has to be extrapolated from the digital simulation in *Maxwell*.

- Phase resistance R_{ph}
- Rotor position θ
- Stator inductances $L_d(\theta)$ and $L_q(\theta)$
- Flux linkages ψ_d and ψ_q

The phase resistance is determined by *Rmxprt* to R_{ph}=16.2$m\Omega$, whereas the rest of the signal data has to be evaluated digitally. The position of the rotor is normally measured by a sensor and can be plotted in *Maxwell*. The values of $L_d(\theta)$ and $L_q(\theta)$ can be plotted during the transient simulation and are depicted in image 9.

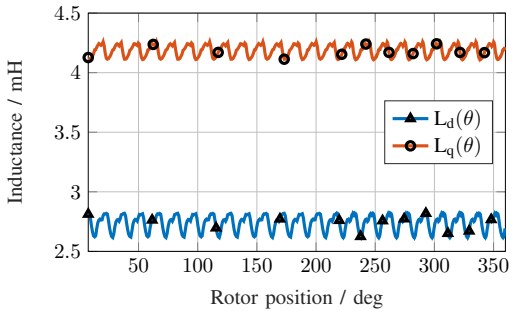


Fig. 9. L_d and L_q over the rotor position at nominal speed of 500 rpm

The average inductance value of L_d is at 2.74 mH, whereas L_q is 4.19 mH. These values depend on the total flux linkage and the current of i_d and i_q.
The values of the phase resistance, L_d and L_q allow to adjust the parameters of PI current controllers for d and q axis currents. These controllers are used to create a q-axis current step of a certain magnitude, producing a torque.
The flux linkage signal of the phases describes the interaction between the coils and the permanent magnets. These are depicted in plot 10.

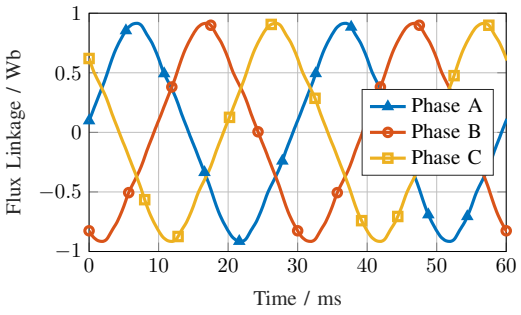


Fig. 10. Flux linkage of the phases A,B and C at nominal speed of 500 rpm

The maximum amplitude of these plots reach $0.915Wb$. This value corresponds to the total flux linkage and describes the sum of the permanent magnetic flux and the product of phase inductance and current. The equation is:

$$\psi_{total} = L_{phase} i_{phase} + \psi_M \tag{10}$$

With the knowledge of these parameters, the fluxes ψ_d and ψ_q can be evaluated by *Maxwell*. In table VII the average values of these quantities are depicted.

TABLE VII
D AND Q FLUX LINKAGE QUANTITIES AT NOMINAL SPEED

Definition	Value	Unit
ψ_d	-0.128	Wb
ψ_q	-0.841	Wb

With the knowledge of the described signal data, a control structure for the motor can be implemented.
It can be concluded that the simulation at nominal speed was successful. It verifies that the motor is able to reach the desired torque in the range of nominal speed up to 500 rpm. The amount of torque ripple lies in an acceptable range for a heavy duty vehicle, but could be reduced by the implementation of an additional torque ripple controller.
Furthermore, it was important to simulate the motor at maximum speed. According to plot 1, the alternative driven diesel engine has to reach a minimum torque of approximately $230N$, at maximum speed of 3100 rpm. At maximum speed, the initial excitation angle is different to the simulation at nominal speed. Another *Optimetrics* simulation led to an initial excitation angle of 22°, in order to achieve maximum torque values. Equally as in the first simulation, a sinusoidal current excitation signal is generated for the 3 phases. The phases are excited with 70 A at a frequency of 206.67 Hz. The simulation time was set to 60 ms, with a sampling time of 0.0001 s. For better visibility of the generated signal curves, only 29 ms of the excitation signals are shown in plot 11.

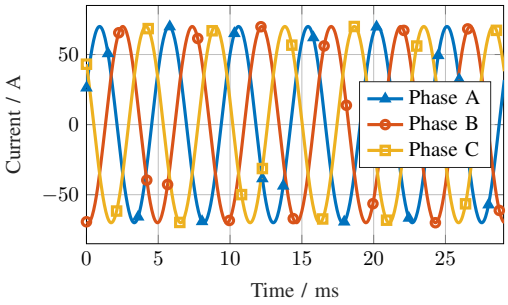


Fig. 11. Input Current signal for 3100 rpm simulation

Again the signal curve is sinusoidal and the phase shifts are distributed correctly. The simulation period of 60 ms corresponds to approximately 3 full revolutions of the rotor.
The simulation at maximum speed led to following output torque signal, depicted in plot 12. For a better signal resolution only 10 ms of the signal is visualized.

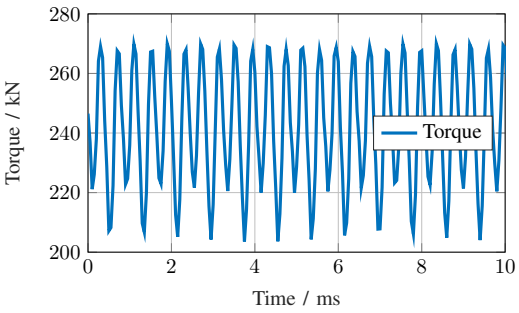


Fig. 12. Output torque signal at 3100 rpm

The average output torque reaches 241.8 Nm. The torque ripple at high speed is approximately ± 30 Nm. This amount of torque ripple is acceptable and won't affect the performance of the electric machine. From the control point of view it is enough to design a controller with the inductance and flux linkage results from the simulation at nominal speed. To this reason the results of L_d,L_q and ψ_d,ψ_q won't be further described.
Another important simulation has to be conducted in the case of no load at maximum speed. The no load simulation is performed by a current excitation of 1 A. This simulation provides information about the permanent magnetic flux linkage and resulting back emf (electro motive force) values. These properties are important parameters, necessary to evaluate the impact on the battery and the inverter for safety reasons [14].

The flux linkage in this simulation varies between -0.52 Wb and +0.52 Wb. It can be understood as the permanent magnetic flux ψ_M during operation and enables to determine the back emf using following equation [3].

$$E_{ph} = p\omega_M \psi_M \tag{11}$$

The resulting signal quantities are shown in following plot 13.

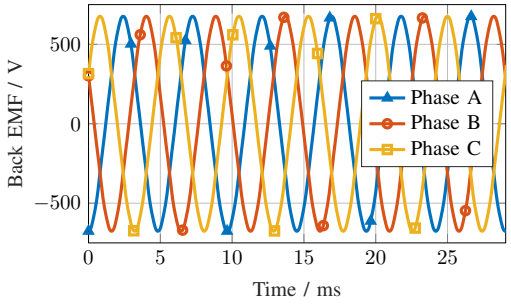


Fig. 13. Emf signal at 3100 rpm at no load

Due to the winding configuration and sinusoidal current excitation the resulting emf signal is also sinusoidal. The back emf reaches a maximum amplitude of 675 V at full speed. This leads to a DC link voltage of

$$V_{DC} = E_{ph}\sqrt{3} = 1169V, \tag{12}$$

exceeding the maximum phase peak voltage of 800 V. This means that the back emf exceeds the power supply voltage of the battery. Without any control structure, the emf signal could damage the battery if the inverter fails. This can occur due to overheating, since the inverter is not able to handle this amount of current. Furthermore, a power supply failure is always a possibility. In order to avoid damage on the battery, a suitable control strategy is mandatory to reduce the DC link voltage. By the control of the amplitude and angle of i_d and i_q, a decreasing voltage results from the i_d current [15].
It can be concluded that the motor reaches the desired parameters of speed, torque and power. The problem of back emf excess can be solved by the implementation of a proper control strategy. Furthermore, it is important to analyse the efficiency of the motor on the basis of power loss calculations and thermal properties. This topic is handled in the next chapter.

VI. Thermal Analysis

In order to investigate the thermal behaviour of the motor a thermal analysis has to be carried out. This also includes an investigation on power losses and the motors efficiency. According to [16], power losses result from resistive losses in the windings, iron losses in the magnetic circuit of stator and rotor and mechanical losses. The losses in the windings occur because of the conductor resistance and can be calculated as

$$P_{Al} = mi^2 R_{phase}. \tag{13}$$

The parameter m stands for the number of phases, i is the stator current and R_{phase}, is the phase resistance. In this equation it gets clear that the power losses in the aluminum winding increase exponentially with the current. These losses increase with the temperature as the conductor resistance gets higher with the operating temperature.
Magnetic losses in the iron circuit result mainly from hysteresis losses and eddy current losses. These losses occur because of magnetization and demagnetization of the core, as a reason of alternating current. Due to a current flow in the winding, the magnetic flux increases up to a point of saturation. Since the iron keeps its magnetization, unless the magnetizing force is applied in the opposite direction (alternating current), this process isn't reversal and leads to power losses. The area included by the magnetic hysteresis loop, indicates the required power for this process, in other words, the power losses. These losses can also be calculated analytically using the Steinmetz formula:

$$P_{hys} = \eta B^n f \tag{14}$$

The quantity η is the material constant, B the peak flux density and f the frequency. The parameter n is an empirical constant and varies between 1.8 and 2.2. As a matter of simplification the eddy current losses are neglected in this project work.
With core loss settings in Maxwell, the losses of the PMSM can be plotted in a transient analysis. These losses were analysed during the simulation at nominal speed. In plot 14 the power loss curves of the rotor and the stator are visualized over 150 ms.

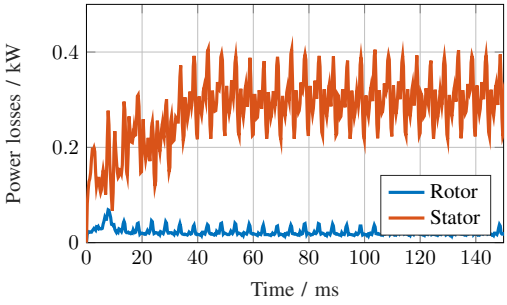


Fig. 14. Power losses in the rotor and stator at nominal speed of 500 rpm

The average loss of the rotor is at 0.02 kW, whereas the average loss in the stator is 0.29 kW. These losses are relatively small, indicating that the magnetic properties of the iron aren't saturated, due to the magnetic flux distribution and temperature distribution.
The main losses during operation occur in the windings, due to the conductor resistance. These losses also rise with temperature, since the material resistivity increases with temperature. In following plot 15, the power losses in the aluminum winding are compared with an identical simulation, using a copper winding, over a simulation period of 300 ms.

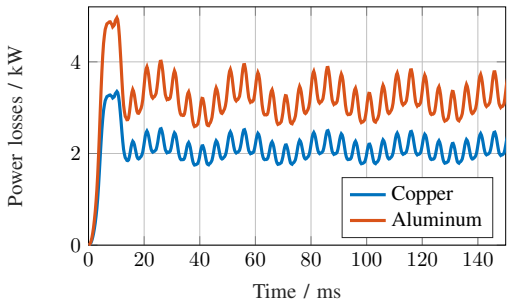


Fig. 15. Power losses in the aluminum winding compared to copper losses at nominal speed of 500 rpm

With an average loss of 3.18 kW, the aluminum losses build the main part of the overall losses. At nominal speed with equal excitation the copper losses are 2.08 kW. The losses with copper are of course lower, since this material has lower resistivity and higher thermal conductivity compared to aluminum [11]. If all the losses are summed up, the motor with aluminum winding looses 3.49 kW at nominal speed. This leads to a motor efficiency of 95.98 %. The motor with copper winding would have an efficiency of 97.2 %, assuming that the stator and rotor losses are equal as in the aluminum configuration.

With the power loss results at nominal speed, a steady state thermal analysis of the PMSM can be conducted. Therefore, it was necessary to develop a coupled simulation between *Ansys Maxwell* and *Ansys Mechanical*. In the *Mechanical* block it is necessary to import the loss data from the 2D Maxwell model. Based on these losses and the material data, the temperature distribution is created.

The main heat evolves in the winding of the motor, due to resistive losses. In order to cool the motor, a heat exchanger can be built around the outer stator surface. In image 16, this construction is depicted for one slot of the winding.

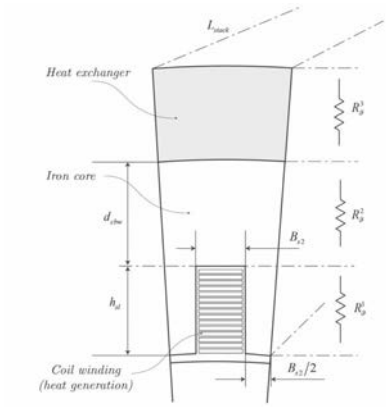


Fig. 16. Thermal model for one slot [3]

This construction can be seen as a thermal circuit. The circuit contains two energy sources, the heat generation from the winding and the cooling from the heat exchanger. The temperature is conducted by thermal resistances and conductors, modelling the inner stator iron, the outer stator iron and the heat exchanger. The resistances depend on the convective heat transfer coefficient λ, in $\frac{W}{m^2K}$. For a correct distribution of the temperature, this quantity can be modelled in *Ansys Mechanical* by the application of convection zones. These convection edges simulate the heat transport in between the motor parts and have to be applied at the outer surfaces of the single parts. The thermal circuit and according formulas are depicted next to the thermal circuit in image 17.

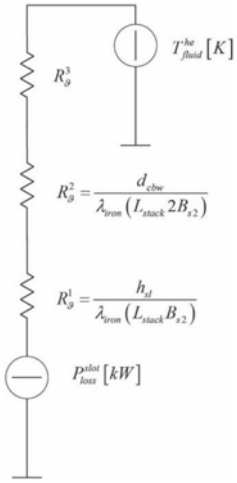


Fig. 17. Thermal model with according thermal circuit of one slot [3]

Since the numerical effort for a thermal transient analysis is immense and goes along with very long simulation times, only a steady state thermal analysis is conducted. The results of this analysis don't deliver information about the heat evolution over time and can't be seen as a fully conducted thermal analysis. Nevertheless, the results give a rough insight in the thermal behaviour of the motor at a steady state. In following table the resulting temperatures are listed for a configuration with copper and aluminum winding.

TABLE VIII
THERMAL ANALYSIS RESULTS OF THE PMSM WITH COPPER AND ALUMINUM WINDING

	Motor part	Min. temperature	Max. temperature
Copper	Stator core	38.8 °C	39.1 °C
	Winding	122.54 °C	122.9 °C
	Rotor core	45.4 °C	45.7 °C
Aluminum	Stator core	42.8 °C	45.7 °C
	Winding	166.40 °C	165.97 °C
	Rotor core	57.36 °C	57.51 °C

The reason for higher temperature results in the copper winding is a higher thermal conductivity and lower electrical resistivity of the material. Since the power losses are lower in

the copper winding this result makes sense.
It has to be mentioned that the whole simulation was carried out in full torque mode. To this reason, the thermal results are upper limit temperature values. Since it is never required to operate a heavy duty machine constantly in full torque mode, the operating temperatures won't be that high as the result testified. On the one hand, this result enables to conduct further, more accurate, analysis in *Maxwell* with adjusted operating temperatures. On the other hand this result goes along with new optimization possibilities. With the conduction of the thermal analysis, the project work of the PMSM design, simulation and analysis could be completed.

VII. Conclusion and future outlook

The presented work in this thesis showed that it was possible to design, simulate and analyze a PMSM model. The iterative design process led to proper torque-speed characteristics in a sufficient range of current and voltage. The basic parameters for control strategies could be evaluated precisely for two different winding configurations with aluminum and copper. Furthermore, it was possible to determine back emf properties of the PMSM model for safety aspects. Besides that, a basic thermal analysis was carried out successfully with a comparison between copper and aluminum winding configurations. The satisfying result opened further optimization possibilities. This goes along with geometrical modifications, in order to reach higher efficiency at lower thermal energy consumption. Therefore, it makes sense to implement a complete control strategy for the motor in a coupled analysis. This is possible using the *Ansys* products of *Maxwell*, together with *Simplorer* or *Matlab/Simulink*. In *Ansys Workbench*, it would be possible to generate a real life operating environment using a coupling with *Mechanical* or *Fluent*. With higher computational power, it would be even possible to implement an extensive transient analysis with coupled 3D PMSM models, similar to a digital twin model. It follows, prototypes of this highly customized motor can be built in a cost saving procedure with low risks as the whole machine was investigated digitally.
Since this motor is operating above the average voltage ranges of $230/400V$, it requires specialized components in the entire power train. The creation of the whole power train goes along with the development of customized components such as the battery or inverters. This could open a wide range of possibilities for mid-sized companies in the future, as the electrification of heavy duty vehicles will be an important topic in the next few years. Since large companies, such as Siemens, might not be interested in the development of these niche products, big opportunities for new start ups can arise.

VIII. Acknowledgment

First of all the authors would like to thank the technical support team of CADFEM for immediate and precise help whenever necessary, regarding $Ansys$. Furthermore, the employees of the Digital Twin Lab of the MCI, especially Mr. Alejandro Secades, have to be mentioned here. Their work built the basis of the whole vehicle electrification project.

References

[1] Gioietta Kuo, "When Fossil Fuels Run Out, What Then?" 2019, MAHB-Millennium Alliance for Humanity and the Biosphere. [Online]. Available: https://mahb.stanford.edu/library-item/fossil-fuels-run/
[2] THOMAS A. LIPO, "INTRODUCTION TO AC MACHINE DESIGN," 2018, Wiley-IEEE Press.
[3] Davide Bagnara, "Electrification of Vehicles," 2021, MCI Management Center Innsbruck.
[4] W.L.Soong, "Sizing of Electrical Machines," 2018, Power Engineering Briefng Note Series.
[5] Without Author, "Electrical Steel: The heart of an electric motor," 2020, [Online; accessed 26.04.2021]. [Online]. Available: https://www.thyssenkrupp.com/en/company/innovation/sustainable-mobility/electrical-steel.html
[6] ——, "NEODYM-EISEN-BOR-MAGNETE," 2019, [accessed 06.04.2021]. [Online]. Available: https://www.arnoldmagnetics.de/products/neodymium-iron-boron-magnets/
[7] Himavarsha Dhulipati, Shruthi Mukundan, "Investigation of Aluminium and Copper Wound PMSM for Direct–drive Electric Vehicle Application ," 2019, Department of Electrical and Computer Engineering, University of Windsor, Canada, IOP Conference Series: Materials Science and Engineering.
[8] J. A. Güemes, A. M. Iraolagoitia, "Comparative Study of PMSM with Integer-slot and Fractional-slot Windings," 2010, XIX International Conference on Electrical Machines - ICEM 2010, Rome.
[9] H. QIU, Y. ZHANG, C. YANG, and R. YI, "Influence of the number of turns on the performance of permanent magnet synchronous motor," 2020, School of Electrical and Information Engineering, Zhengzhou University of Light Industry, Zhengzhou, 450002, China, pp. 429-432.
[10] Without Author, "Trading economics, Commodities Prices," 2021, [accessed 20.07.2021]. [Online]. Available: https://tradingeconomics.com/commodities
[11] ——, "Kupfer in der Elektrotechnik - Kabel und Leitungen," 2000, Deutsches Kupfer-institut e.V, p.5 .
[12] Mario Nikowitz, Matthias Hofer, Manfred Schroedl, "Reduction of the 6th and 12th harmonic in the torque ripple of a salient pole synchronous reluctance machine ," 2019, Institute of Energy Systems and Electrical Drives, Technische Universitaet Wien, Vienna, Austria.
[13] Joachim Holtz, Lothar Springbob, "Identification and Compensation of Torque Ripple in High-Precision Permanent Magnet Motor Drives ," 1996, IEEE TRANSACTIONS ON INDUSTRIAL ELECTRONICS, VOL. 43, NO. 2, .
[14] Without Author, "Overspeed, back-EMF and rectified voltage," 2021, [accessed 27.07.2021]. [Online]. Available: https://danfosseditron.zendesk.com/hc/en-gb/articles/360018207257-Overspeed-back-EMF-and-rectified-voltage
[15] Tim Martin, Richard Burke, "Practical Field Weakening Current Vector Control Calculations for PMSM in Vehicle Applications ," 2013, EVS27 International Battery, Hybrid and Fuel Cell Electric Vehicle Symposium, Barcelona, Spain .
[16] Co Huynh,Liping Zheng,Dipjyoti Acharya, "Losses in High Speed Permanent Magnet Machines Used in Microturbine Applications ," 2009, Journal of Engineering for Gas Turbines and Power, Vol 131, Calnetix, Inc., Cerritos, CA 90703.

Lukas Nuschele is currently enrolled in the Masters program "Mechatronics and Smart Technologies" at the Management Center Innsbruck. He has a bachelors degree in mechanical engineering and successfully finished an apprenticeship as electronics for automation technology at DMG Mori.

Edge-AI State Classification using a Supervised-Deep-Learning CNN with an Automated Measurement System Interface

Philipp Roch and Martin Pillei (supervisor)

***Abstract*—To ensure the profitability of processes at diverse operating systems, down times must be minimized, efficiency increased, and a high quality level ensured. Thus, implementing an automated and universal condition monitoring (CM) and predictive maintenance (PM) system is necessary. Hence, the Edge-Cloud-AI system is presented focusing the on the edge artificial intelligence (AI) and its deep learning (DL) implementation regarding universal applicability. The innovative kind of CM and PM by the presented Edge-Cloud-AI prototype system is the real-time prediction of the current operating state on the edge, detecting changing operating states and variations during the process. To achieve this, a modular system is established controlled by a graphical user interface (GUI) considering data acquisition, data preparation, neural network (NN) training, and real-time state classification by the developed and trained NN model. The modular software architecture and the measurement device ensured the evaluation by a preliminary test application incorporating 1D and 2D NN model classification methods. The focus is set to match the complexity of the data while achieving the highest efficiency at the real-time state classification by tailored workflows of software and hardware. By the use of ordinary and manipulated test datasets, the implemented NN methods for state classification are evaluated. As a result, convolutional neural networks (CNNs) as 2D model and in combination with long short term memory as 1D model exhibit a high accuracy for the investigates application at a sufficient efficiency. Further trial runs with the functional prototype approved the established workflow to be easy to use, efficient, and highly functional. This indicates that the Edge-Cloud-AI system can be a suitable, innovative alternative for a multitude of application. Further tests of more complex real applications by the finalized prototype are yet to be conducted.**

***Index Terms*—Predictive Maintenance, Deep Learning, CNN, Classification, Edge-AI, Fluid Flow**

P. Roch is with the Department of Mechatronics, MCI, Austria, e-mail: ph.roch@mcie4me.at and M. Pillei is with the Department of Industrial Engineering, MCI, Innsbruck, Austria E-mail: martin.pillei@mci.edu.

I. Introduction

REAL-time CM and PM enables the monitoring of conditions and states in a certain process or manufacturing environment. Down times and less efficient production processes are the result of undetected disturbances and failures in the processing circle that are not different to certain patterns during the operation. [1] These patterns differentiate at different process states and can be detected by various 1D and 2D sensors. DL algorithms are effectively able to learn patterns extracted from the data gathered by the sensors and generalize to the system in order to predict a certain state that is currently operating at the process or the machine. [2], [3] Thus, classification of conditions and states is not trivial but necessary to get the most out of a complex system. [4]

The motivation for the research and development presented in this thesis, is the Edge-Cloud-AI system as a universal application for CM and further for PM independent of the system or process. The system is a extension of the system proposed by [1], [5], and [6], including the deployment of AI at the edge according to [7] and [8]. This will give a real-time insight to what state the machine is operating over the time of use, what states or changes in the operation process caused failures, and what measures

and actions must be taken. [9] The aspects of on the edge data processing, cloud computation, and intelligence directly at the process can set the trend for future CM and PM. More precisely, the future of data driven CM and PM is based on anomaly and pattern detection by DL models. [10]

II. Methods

Within the research and development ensuing methods are split in system development, measurement setup, and the NN design. Several subtasks are to be distinguished in order to match the requirements of the system.

1) Data acquisition
2) Modular sensor selection
3) Variable batch sizes
4) NN training
5) Edge-Cloud interface
6) State classification
7) State visualization

The presented Edge-Cloud-AI system comprises of all stated operations. Hereby, hardware and software are the main parts and thus considered for the software architecture. The accordingly developed measurement device is shown in figure 1.

Fig. 1. Measurement Device

A. System Architecture

The system components are displayed divided in edge and cloud components. The graphical user interface (GUI) and the measurement device describe the Edge-AI part, whereas the server defines the cloud component of the developed system. The high level software structure including all main components implemented at the measurement device is depicted in figure 2.

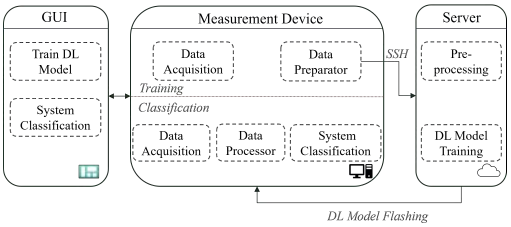


Fig. 2. High level system overview

The system architecture ensures the capability of the system to operate right on the edge alongside the computational power of the server for the training of the implemented 1D and 2D NNs. The GUI provides the front-end interface for inputs and outputs alongside according back-end functions to control the entire system. All functions despite parts of the data preprocessing and the NN training are executed by the system-on-a-chip (SoC) and as stated headed by the GUI.

B. Hardware

The inside of the device comprises of the raspberry pi SoC, a breadboard and its analog-digital converter (ADC) for prototyping, six printed circuit boards (PCB) for the connection of the sensors by JST-XH connectors, and a radiator for device cooling. A simplified schematic of the inside structure is depicted in figure 3.

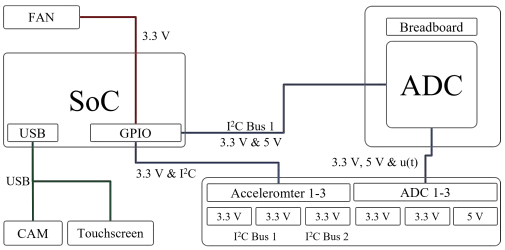


Fig. 3. Hardware schematic of the measurement device

C. Software

For the software a system of several scripts and components is developed in python. Additionally, system files for the SSH interface to the server are implemented and controlled by the main python scripts. The basic flow of the developed software architecture is displayed in figure 4.

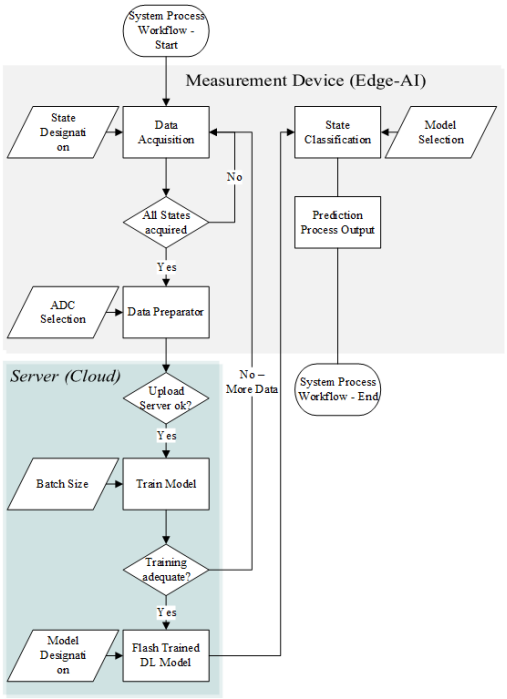


Fig. 4. Flow chart of the software architecture

First, the data acquisition is executed. With all desired states gather and processed, the data preparator concatenates the states by their labels, removes non-selected or non-detected input features and uploads the datasets to the server. The modular architecture of the system allows the user, additionally to the regular 1D and 2D sensors, to stand-alone train the algorithm by the 1D sensors or the 2D sensors. On the server, the NN model is trained with defined parameters and sensors and afterwards flashed to the measurement device. Here, the SoC executes the state classification by the flashed NN models.

D. Measurement Framework

The measurement setup simulates a minimal working example of the overall system. Therefore, measured and generated data are designed to be learned and differentiated well. Further, by setting up a proper real-time prediction setup, the system is validated in order to work for more complex systems. The real-time prediction setup consists of unseen data extracted of the datasets before training and additional manipulations at several sequences.

The test rig for the real-data measurement data to validate the 1D NN model is depicted in figure 5. According sensors are marked as multi-axial accelerometer, pressure, and sound sensor.

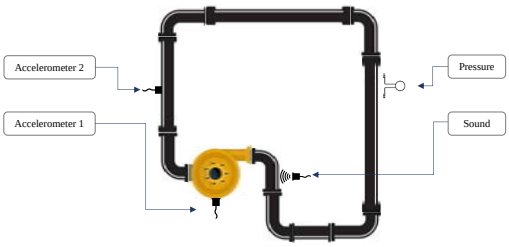


Fig. 5. Schematic of the measurement setup

In order to validate the 2D NN model by an image classification dataset, generic flow simulation data is used considering varying velocities v.

E. Data Preprocessing

As the data acquired by the measurement device is not normalized, labeled, split, or shuffled, several steps are mandatory to feed the data to the NNs. Therefore, the first step is the data preparation on the SoC. Further, the labeling, shaping, shuffling, and splitting of the uploaded dataset is carried out directly

at the NN model training script. The data split is defined as:

- Training data: 0.72
- Validation data: 0.18
- Testing data: 0.1

F. Neural Network

Within the context of NNs, two different approaches are considered to be suitable architectures. For the time-series classification (TSC) a custom 1D CNN is designed, whereas for the image classification (IC) both a custom 2D CNN and pretrained models are designed. The method of transfer learning is implemented by the pretrained models of the Keras library. [11] The procedure for training implemented NN models is shown in figure 6.

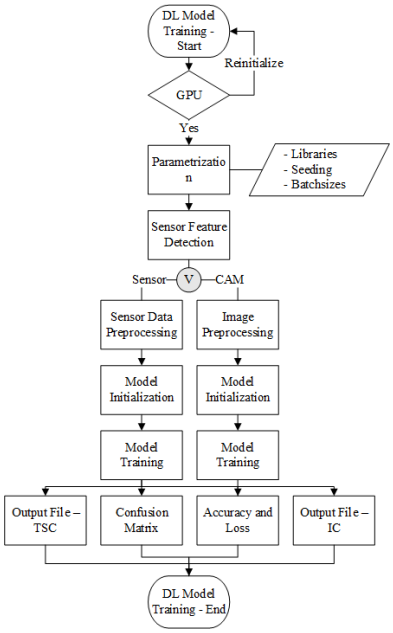


Fig. 6. Flow chart of the NN training process

All considered models for TSC and IC are adjusted not to over-, or underfit and increase the ability to generalize to unseen data. Further, the hyperparameters are adjusted according to the grid search method. [3], [12], [13]

For the TSC, the best performing FCN - LSTM architecture is implemented according to [14] and [15]. It unites the feature extraction for time-series of fully convolutional network (FCN) and the capability of the long short term memory (LSTM) to memorize what features are important at the long-term. [16], [13], [3]
The preprocessed segments vector $\vec{x}$ for the data sequences is fed to the NN as a 2D vector and reshaped by a reshape layer to a 3D vector comprising of the segment length and the number of features. In this context, a multivariate model with 8 features. Subsequent, the data sequence is fed to two dense layers with a dropout layer in between. The fully connected layers of filter size 128 and 64 neurons define a first neuron sequence that matches the input shape size besides the dropout for deactivating insignificant neurons. [12] Afterwards, the model splits to a FCN feature extraction line and a LSTM sequence with two preceded 1D convolutional layers for feature extraction. All 1D convolutional layers are activated by the rectified linear unit (ReLU) activation function to detect nonlinear patterns at a sufficient efficiency. [3], [12] The FCN line consists of three blocks of 1D convolutional layers, batch normalization and a ReLU activation function. Filter sizes narrow down to 32 in both the FCN and the LSTM line ensuring a sufficient complex feature extraction by not overfitting the NN model. [14], [15] Furthermore, the kernel size for each feature map regarding the first 1D convolutional layer at the FCN line is set to 50 in order to start with a good range for pattern detection. Advancing in the NN deepness, the kernel size decreases to detect more detailed patterns. [3] Implemented layers like dropout and batch normalization prevent the model from overfitting and improve the regularization process. [17], [12] At last, the FCN feature extraction line is global average

pooled to get a 2D vector shape and subsequent, both feature extraction lines are concatenated and fed to the final dense layer outputting the shape of defined states. Hereby, the softmax activation function calculates the neurons activation to an absolute range of 0 to 1. Adequate hyperparameters of the FCN - LSTM are defined as follows:

- Epochs: a maxima of 50 epochs to prevent overfitting
- Batch size: 10 segments at a time
- Learning rate and optimizer: 0.0001 (adam optimizer)

For the IC, the best performing 2D CNN architecture is implemented according to [18], [19], [2] to match the complexity of the dataset. More complex structures stated as the XCeption, the ResNet, or the EfficientNet are considered at the evaluation, but exceeding the sufficient NN complexity for the data fed.

The preprocessed segments vector $\vec{x}$ for the data sequences is fed to the NN as a 4D vector and the pixel values normalized by a rescale layer. Subsequent, the vector $\vec{x}$ is fed to three consecutive blocks consisting each of a 2D convolutional, a dropout, and a 2D maxpooling layer. All 2D convolutional layers are activated by the relu function to detect nonlinear patterns at a sufficient efficiency. [3], [12] To mention, dropout and maxpooling layers are added to each block to prevent overfitting and increase the regularization process. Filter and kernel sizes of implemented 2D convolutional layers range from 16 up to 64 and from 5 down to 3. The increased filter size up to 64 at a narrowed kernel size allows the NN to detect a higher variety of features at a more complex level. [3], [12] Simultaneous, the increase of the dropout ensures the deactivation of less significant neurons at categorized features. At last, the output shape of the last block is flattened to get a 2D vector shape and fed to a final dense layers. Hereby, the softmax activation function of the last calculates the neurons activation to an absolute range of 0 to 1. Adequate hyperparameters of the 2D CNN are defined as follows:

- Epochs: a maxima of 50 epochs to prevent overfitting
- Batch size: 5 2D images at a time
- Learning rate and optimizer: 0.001 (adam optimizer)

The procedure for the real-time state classification by sufficiently trained NN models is shown in figure 7. Here, the NNs output is reviewed and analyzed by a voter outputting the best real-time state classification.

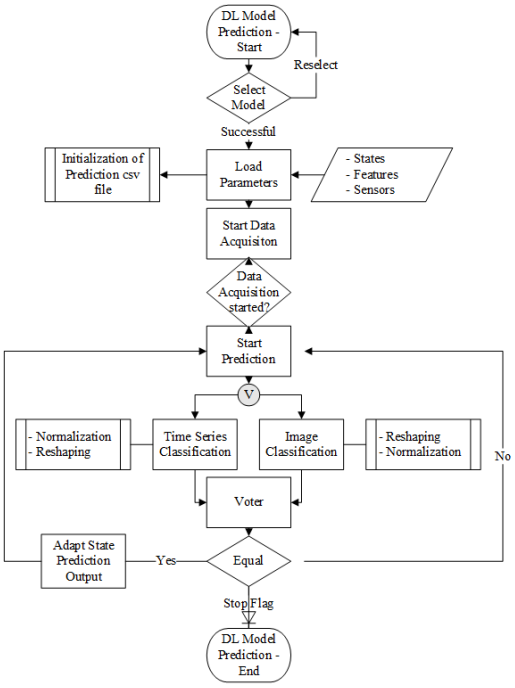


Fig. 7. Flow chart of the NN real-time state classification process

III. Results

To evaluate the distinguished NN architectures by gathered 1D and 2D data, measurement data is

presented and particular metrics analyzed and summarized by the performance-efficiency index (PEI) PEI1 and PEI2.
The F1 score defines as a measure of accuracy as stated in [3]. Further, the inference step time t_i, the prediction step time t_p, the standard deviation σ, the NN parameters, and most important the real-time classification (RT) score as a metric of classification and generalization ability of the NN to unseen and unseen manipulated data are considered. Here, the data is split to 2 categories by their degree of manipulation. Subsequent, stated metrics are computed to weighted sums defining the PEI1 and PEI2. The F1 score is calculated by the precision p and the recall r according to

$$F_1 = \frac{2pr}{p+r} \tag{1}$$

and the RT score according to

$$\begin{aligned} SCORE_{RT} = \\ (N_{+,C_0} \cdot 1 + N_{+,manipulated,C_1} \cdot 3 \\ + N_{+,manipulated,C_2} \cdot 5 \\ + \sum N_{+,C,P<0.9} \cdot P_+) \\ -(N_- \cdot 5 + N_{-,manipulated} \cdot 3) \end{aligned} \tag{2}$$

where

N ... the amount of right $(+)$ and wrong $(-)$ predictions
P ... the probability in $\%$
C ... the manipulation category.

A. Measurement Results

For the 1D TSC, the normalized data sequences of the four defined states regarding the the multi-axial accelerometers, the pressure sensor, and the sound sensor are acquired. The multi-axial accelerometer data is plotted in figure 8 displaying states 1 to 4 from sample 0 to 2×10^6.

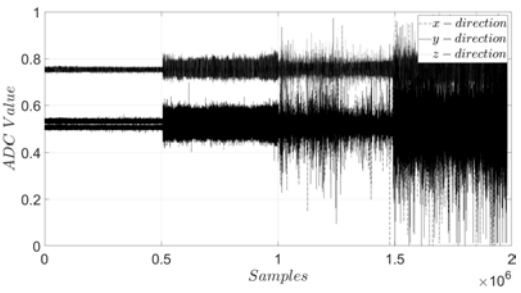


Fig. 8. Multi-axial accelerometer data - state 1 (left) to 4 (right)

For the 2D IC, two sample plots for the lowest and highest velocity v are shown in figure 9.

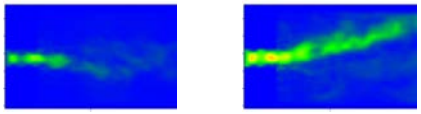


Fig. 9. Flow at various velocities v - $0.4\frac{m}{s}$ (left) and $1\frac{m}{s}$ (right)

For both the 1D and 2D data clear differentiations between the states are exposed. In figure 10 the sum of 1D and 2D data gathered for each state and the overall balance of the dataset is visualized.

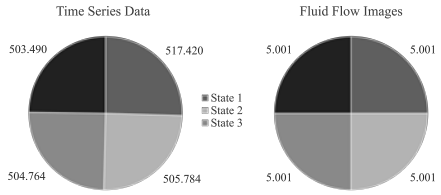


Fig. 10. Data split of the used datasets

B. Neural Network Performance

The performance metrics of the most suitable FCN - LSTM and the 2D CNN model show great progress

over the 47 or 49 trained epochs. The defined batch size of 10 segments at the time-series and 5 at the 2D images trained in parallel show great convergence to the training and validation dataset. The NN models performance assessment by the validation loss at the validation dataset shows equally great convergence towards the 47^{th} and respectively 49^{th} epoch of training. Hence, the models are not overfitting nor underfitting, resulting in a good regularization process and therefore good generalization towards unseen data. The training and validation process described is shown in figure 11 for the TSC and in figure 12 for the IC.

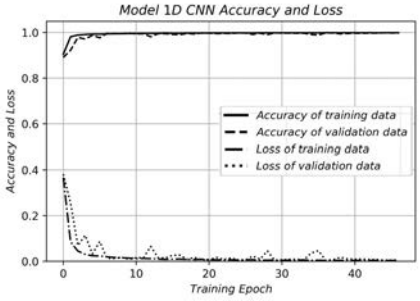


Fig. 11. Model parameters of the FCN - LSTM - Performance on the testing dataset

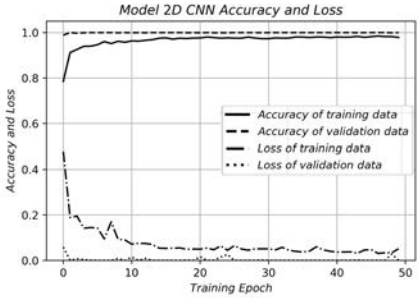


Fig. 12. Model parameters of the 2D CNN - Performance on the testing dataset

The indicated ability to generalize and therefore to perform well at unseen data is approved by the testing dataset. The output is depicted in a confusion matrix in figure 13 for the TSC and 14 for the IC, both stating an outstanding accuracy of $99.5\,\%$ and $100\,\%$ at unseen data by comparing the true label with its according predicted label.

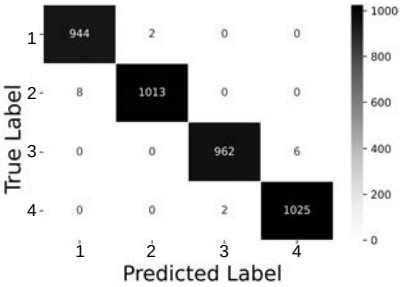


Fig. 13. Confusion matrix of the FCN - LSTM - Performance on the testing dataset

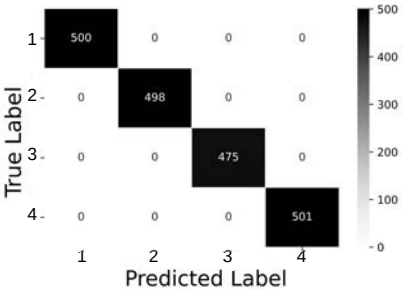


Fig. 14. Confusion matrix of the 2D CNN - Performance on the testing dataset

The described PEI summarizes the characteristics of all evaluated NN models and indicates the best suitable to be implemented in the final system. PEI1 considers the NN model metrics for training, evaluation, and real-time prediction, whereas PEI2 describes the efficiency of each NN model.

Hence, the PEI1 and PEI2 for all compared NN models are depicted in table I and II. The smaller the PEI score, the better is the suitability for the developed system. The real-time prediction (RT) and F1 score are hereby computed according to equation 1 and2.

TABLE I
PERFORMANCE EFFICIENCY INDEX 1 (PEI1)

NN model	σ	F1 score	RT score	PEI1
1D CNN	$6.732e^{-3}$	0.992	67.47	0.091
FCN - LSTM	$4.452e^{-3}$	0.984	67.58	**0.088**
CNN - LSTM	$6.644e^{-3}$	0.994	57.35	0.105
1D ResNet	$2.361e^{-3}$	0.995	62.08	0.925
2D CNN	$3.273e^{-3}$	1.000	58.95	**0.089**
EfficientNetB2	$4.461e^{-3}$	1.000	58.65	0.091
ResNet50V2	$7.199e^{-3}$	1.000	58.55	0.097
XCeption	$5.048e^{-3}$	1.000	54.10	0.996

TABLE II
PERFORMANCE EFFICIENCY INDEX 2 (PEI2)

NN model	NN size in MB	Inference step GPU t_I / ms	Prediction step SoC t_p / $ms \pm \sigma$	PEI2
1D CNN	2.932	7	212 ± 22	**0.50**
FCN - LSTM	3.681	13	558 ± 13	1.35
CNN - LSTM	3.071	64	842 ± 54	2.15
1D ResNet	21.110	15	656 ± 8	2.01
2D CNN	28.732	9	956 ± 98	**0.68**
EfficientNetB2	92.195	113	358 ± 98	1.29
ResNet50V2	276.811	75	1859 ± 146	2.50
XCeption	244.897	150	1420 ± 162	2.65

As it can be seen, the previous described best suitable NN models outperform residual ones by the most important RT score at a sufficient efficiency.

C. System Results

To conclude, the developed software system works as expected. GUI, data acquisition by controlling all connected sensors, server interface, model training, and real-time state classification are fully functional and perform without failures. The implemented modular architecture ensures the system to process data dependent on the system states and necessary sensors. Further, the system is able to operate stand-alone by 1D or 2D sensors.

Considering the performance of the NN, the PEI 1 goes in line with the theoretical research. The accordingly designed NN models perform as desired and achieve $> 99\,\%$ at the testing set and a RT score percentage of $> 90\,\%$ including highly manipulated and varying data. The FCN - LSTM for TSC is optimized and adjusted for both the FCN and the LSTM feature extraction line. For the FCN, optimizations especially for the filter and kernel size proofed to be significant, whereas at the LSTM feature extraction line, the influence of the before extracted feature by the preceded convolutional layers showed valuable improvements.
Regarding the 2D CNN, adjustments to the kernel, the filter size, and the depth of the NN showed a significant improvement in generalization ability. Rather simple structures outperform more complex NN architecture indicating the matching complexity of the 2D CNN in contrast to pretrained and more complex NN models. In general, due to the grid search hyperparameter optimization and adjustments to the NNs complexities, the desired generalization and invariance in terms of reproducibility at varying datasets of the NN is achieved, ensuring the applicability of the software system to other systems.

The developed measurement device including the SoC, the ADC, the sensors and its connectors, the input device and the server architecture work as expected. The hardware limited sampling rate works sufficiently well and achieves the desired sampling rate throughout the process flow of $200\,Hz$ to $250\,Hz$. The PEI 2 states that storage and performance of the SoC are adequate to process and display data, build the interface to the server,

and to real-time predict the current state of the operating system. The prediction step time of the SoC ranges from $558\,ms$ for the TSC up to $956\,ms$ for the IC and is therefore suitable to the interval the system is predicting the operating state. Considering the performance of the server, needed resources are given and the training executed on the GPU. As a result, the inference step of the training process amounts to $13\,ms$ for the FCN - LSTM and to $9\,ms$ for the 2D CNN. The training process can therefore be processed sufficiently well.

IV. Conclusion and Outlook

The research and the presented results considering this thesis define an auspicious prospect for the future of CM and PM. The developed system demonstrates a prototype of the Edge-Cloud-AI system for the particular and universal use at production sites or laboratories. The innovative kind of CM and PM by the presented system is the real-time prediction of the current operating state on the edge, detecting changing operating states and variations during the process. To achieve this, a modular system is established controlled by a GUI considering data acquisition, data preparation, NN training, and real-time state classification by the developed and trained NN model. The modular software architecture and the measurement device ensured the evaluation by a preliminary test application incorporating 1D and 2D NN model classification methods. The focus is set to match the complexity of the data while achieving the highest efficiency at the real-time state classification by tailored workflows of software and hardware. By the use of ordinary and manipulated test datasets, the implemented NN methods for the real-time state classification are evaluated. Conclusive, CNNs as 2D NN models and in combination with LSTMs as 1D NN models exhibit a high accuracy for the investigated application at a sufficient efficiency. Further trial runs with the functional prototype approved the established workflow to be easy to use, efficient, and highly functional. The critical review of the simplifications, assumptions, and the general proof of concept clarifies that the developed Edge-Cloud-AI system can be stated as a cost efficient system real-time classifying the current operating state of a particular learned system at a high performance. The base for the application as a universal CM and PM device is elaborated, thus the functional prototype can be considered to as an innovative alternative to state of the art approaches and processes. In order to optimize and extend the functional prototype Edge-Cloud-AI system to a standardized, industrial product, several areas of improvement are stated that are not specifically considered in more detail at the presented research

Stated as SoC and processing hardware, data and preprocessing, NN enhancement, and the general approach of DL, several areas of improvement are considered. Dependent on the application, the importance of each area varies.

Acknowledgment

The author would like to thank Martin Pillei and Manuel Berger for the great research opportunity at the research cluster of fluids and mechanics. Constructive meetings, support, and motivation throughout the thesis guided the research and development to a better result.

References

[1] W. Sun, J. Liu, and Y. Yue, “Ai-enhanced offloading in edge computing: When machine learning meets industrial iot,” *IEEE Network*, vol. 33, pp. 68–74, 2019.

[2] Y. Lecun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, pp. 436–444, 2015.

[3] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016. [Online]. Available: https://www.deeplearningbook.org/

[4] A. P. Kane, A. S. Kore, A. N. Khandale, S. S. Nigade, and P. P. Joshi, “Predictive maintenance using machine learning,” 2022.

[5] S. Kum, Y. Kim, and J. Moon, “Deploying deep neural network on edge-cloud environment,” *ICTC 2019 - 10th International Conference on ICT Convergence: ICT Convergence Leading the Autonomous Future*, pp. 242–244, 2019.

[6] K. S. Kiangala and Z. Wang, "Initiating predictive maintenance for a conveyor motor in a bottling plant using industry 4.0 concepts," 2018.
[7] A. M. Ghosh and K. Grolinger, "Edge-cloud computing for internet of things data analytics: Embedding intelligence in the edge with deep learning," *IEEE Transactions on Industrial Informatics*, vol. 17, pp. 2191–2200, 2021.
[8] X. Wang, Y. Han, V. C. M. Leung, D. Niyato, X. Yan, and X. Chen, *Edge AI*. Springer Singapore, 2020.
[9] T. Yeung, "What is edge ai and how does it work?" 2022. [Online]. Available: https://blogs.nvidia.com/blog/2022/02/17/what-is-edge-ai/?ncid=so-link-326670&=&linkId=100000119999863#cid=dl23_so-link_en-us
[10] G. Aydemir and B. Acar, "Anomaly monitoring improves remaining useful life estimation of industrial machinery," *Journal of Manufacturing Systems*, vol. 56, pp. 463–469, 2020.
[11] Keras, "Keras layers api." [Online]. Available: https://keras.io/api/layers/
[12] A. Zhang, Z. C. Lipton, M. Li, and A. J. Smola, *Dive into Deep Learning*, release 1.0.0-alpha0 ed., 2021.
[13] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, "Backpropagation applied to handwritten zip code recognition," *Neural Computation*, vol. 1, pp. 541–551, 1989.
[14] F. Karim, S. Majumdar, and H. Darabi, "Lstm fully convolutional networks for time series classification," 2017.
[15] F. Karim, S. Majumdar, H. Darabi, and S. Harford, "Multivariate lstm-fcns for time series classification," 2019.
[16] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, pp. 1735–1780, 1997.
[17] B. Mehlig, *Machine learning with neural networks*, 2021.
[18] A. Ajit, K. Acharya, and A. Samanta, "A review of convolutional neural networks," *International Conference on Emerging Trends in Information Technology and Engineering, ic-ETITE 2020*, 2020.
[19] S. Albawi, T. A. Mohammed, and S. Al-Zawi, "Understanding of a convolutional neural network," *Proceedings of 2017 International Conference on Engineering and Technology, ICET 2017*, vol. 2018-January, pp. 1–6, 2018.

Philipp Roch is an engineering master graduate student at the department of mechatronics and smart technologies at the Management Center Innsbruck (MCI).

Control Concept for Gas Engines with Stoichiometric Combustion

Michael Schmid, Johannes Huber (supervisor), and Andreas Mehrle (supervisor)

***Abstract*—Future regulations on modern gas engine power plants enforce a further reduction of already low exhaust gas emissions, such as NO_x, CO and HC. This requires novel approaches in engine control methods and exhaust aftertreatment. Most large bore gas engines are operated in a lean burn operation principle to meet emission regulations. Here an alternative approach is investigated, which relies on a stoichiometric combustion principle.**

The work focusses on optimising the control of the system of engine and exhaust aftertreatment unit - the three-way catalyst (TWC). Different control methods are developed using a catalyst model that reproduces the oxygen storage capability of a TWC. Performance and robustness criteria are compared for an H_∞ robust control and a classical cascade PID control method. Catalyst parameters are further estimated online using adaptive control techniques.

The developed control concepts allow stable operation of the engine up to full load and ensure, that the catalyst is kept in a favourable range of operation. Thereby tailpipe exhaust emission targets can be fulfilled and the ability to comply with future emission regulations is demonstrated.

***Index Terms*—Large bore gas engine, stoichiometric combustion, three-way catalyst, robust control, parameter estimation.**

I. Introduction

THE precise control of the air-fuel ratio AFR is important in modern engine control systems. Not only does it manipulate the rather delicate quantity of fuel injected into the engines combustion chamber, it also plays a distinct role in exhaust gas emissions. The regulations on the latter are becoming more and more stringent [1], which results in lower tolerable margins for control deviation. On the other hand, catalysts contain noble metals, which are getting increasingly expensive due to ceasing natural resources and in turn leads to tight dimensioning of components with little reserve and more fragile constructions. Despite those challenging boundary conditions, engine manufacturers must guarantee stable operation of their products to ensure customer satisfaction and are interested in low service costs and simple diagnostics.

These aspects all influence the design of a controller and especially the requirements on its robustness and reliability. To this end, a H_∞ robust controller should be designed to control the AFR of a large bore gas engine equipped with a three-way catalyst (TWC). The operating point of the TWC must be kept in a tight range, where its conversion efficiency is high and exhaust gas emissions can be effectively reduced. Different approaches have been reported for efficient control of the air-fuel ratio of an engine equipped with a three-way catalyst. Schallock et al. [2] demonstrate model predictive control (MPC) with a relay controller, Trimboli et al. [3] describe an MPC approach with delay compensation, Shafai et al. [4] report online system identification of fuel path time delay and Ingram et al. online system parameter estimation of the catalyst oxygen storage capacity [5]. This work is motivated by the demonstration of a robust H_∞ control method by Roduner et al. [6] and subsequent work by collaborators on adaptive control methods, Shafai et al. [7] and Ammann et al. [8].

The paper is organised as follows. In Section II the used catalyst model, H_∞ controller design methods and TWC parameter estimation are presented. Section III demonstrates measurements and simulation results of AFR control and adaptive control for parameter estimation. Conclusions are summarized in Section IV.

M. Schmid is with the Department of Mechatronics, MCI, Innsbruck, Austria, e-mail: sm1211@mci4me.at.

II. Methods

A. Overview

The key quantity to be controlled is the air-fuel ratio AFR. It is defined via the mass flows of air $\dot{m}_{\mathrm{air}}$ and fuel $\dot{m}_{\mathrm{fuel}}$ (in the considered case natural gas) entering the combustion chamber of a spark ignited reciprocating engine,

$$AFR = \frac{\dot{m}_{\mathrm{air}}}{\dot{m}_{\mathrm{fuel}}}. \tag{1}$$

A common derived quantity, widely used in engine control language, is the lambda coefficient, also called air-fuel equivalence ratio. It describes the ratio of the present AFR with the air-fuel ratio required for stoichiometric combustion $L_{\min}$,

$$\lambda = \frac{AFR}{L_{\min}}. \tag{2}$$

The latter depends on the type of fuel used. If an engine is operated with excess air, it follows the so called lean burn principle with $\lambda > 1$. In contrast, if the combustion takes place with air deficit, one speaks of rich operation with $\lambda < 1$. At stoichiometry $\lambda = 1$, which is often referred to as rich burn principle.

B. Fuel Path Approximation

Efficient AFR control of a reciprocating engine requires a model of the fuel path, where the section from gas injection valve up to TWC inlet needs to be considered. The fuel flow is governed by a gas mixing process followed by a substantial transport delay through the intake manifold, combustion chamber and exhaust pipe. This structure is schematically illustrated in Fig. 1.

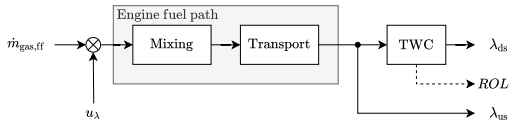


Fig. 1: Simplified fuel path model of a reciprocating engine with a three-way catalyst.

The actuating variable of the AFR controller is implemented as a multiplicative factor u_λ on the nominal amount of gas injected in the fuel path $\dot{m}_{\mathrm{gas,ff}}$. The latter is calculated from an estimated mass flow composition into the engine as well as a desired $\lambda_{\mathrm{us,set}}$. Measurements of lambda upstream λ_{us} and downstream λ_{ds} the catalyst are used for feedback and the relative oxygen storage level ROL in the TWC is critical for its conversion efficiency.
Modifying the gas dosage influences the λ value, according to Eqs. (1) and (2). Taking into account the mixing and transport delay processes and linearising around the nominal value of $u_\lambda = 1$, a plant model for the fuel path is found [6]. Using $\Delta u_\lambda = u_\lambda - 1$ and $\Delta\lambda_{\mathrm{us}} = \lambda_{\mathrm{us}} - \lambda_{\mathrm{us,set}}$, the influence of Δu_λ on $\Delta\lambda_{\mathrm{us}}$ is expressed as

$$G_\lambda = \frac{\Delta\lambda_{\mathrm{us}}}{\Delta u_\lambda} \approx \frac{K_{\mathrm{fp}}}{s\tau + 1} e^{-\delta s}, \tag{3}$$

with K_{fp} being the proportionality constant, τ the mixing time constant and δ the transport delay time. These parameters depend on the working point of the engine, i.e. on speed and power.

C. Three-way Catalyst

1) Modelling: To cover the complexity of reduction and oxidation processes in a TWC, a system of coupled rate equations is often used for detailed investigations [9], [10]. However, for control purposes such models would require too high computational effort to be used effectively, which stresses the need for a simplified control-oriented model of the TWC dynamics.
A TWCs main characteristic is the ability to store and release oxygen. For control purposes it suffices to reproduce this property. The implementation by Brandt et al. [11], [12] models the relative oxygen storage level ROL of the TWC as in Eq. (4). The storage level is limited to a range of $ROL \in [0, 1]$, with 0 corresponding to a fully depleted and 1 to a completely filled O_2 storage,

$$\dot{ROL} = \begin{cases} 0.23\ \dot{m}_{\mathrm{exh}} \frac{\rho(\lambda_{\mathrm{us}}, ROL)}{C_{\mathrm{cat}}} \left(1 - \frac{1}{\lambda_{\mathrm{us}}}\right) & 0 \le ROL \le 1, \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

Here $\dot{m}_{\mathrm{exh}}$ denotes the exhaust gas mass flow entering the catalyst, λ_{us} is the lambda upstream catalyst and C_{cat} the oxygen storage capacity. The factor 0.23 is the oxygen mass fraction of air. Oxygen exchange of the exhaust gas with the catalyst material is influenced by the empirical function ρ, which reproduces the typically different O_2 storage and release rates of a TWC. Equation (5) indicates the dependency on the current ROL as well as the exhaust gas being in the lean ($\lambda > 1$) or rich ($\lambda < 1$) range,

$$\rho(\lambda_{\mathrm{us}}, ROL) = \begin{cases} \alpha_{\mathrm{L}} f_{\mathrm{L}}(ROL) & \lambda_{\mathrm{us}} \ge 1, \\ \alpha_{\mathrm{R}} f_{\mathrm{R}}(ROL) & \lambda_{\mathrm{us}} < 1. \end{cases} \tag{5}$$

The intensity of storage and release rates can vary, which is covered by different α_{L} and α_{R} factors. Furthermore, the probability functions for storage and release of an oxygen molecule, f_{L} and f_{R}, are defined in Eqs. (6) and (7),

$$f_{\mathrm{L}}(ROL) = 1 - \frac{e^{\phi_{\mathrm{L}} ROL} - 1}{e^{\phi_{\mathrm{L}}} - 1}, \tag{6}$$

$$f_{\mathrm{R}}(ROL) = \frac{e^{-\phi_{\mathrm{R}} ROL} - 1}{e^{-\phi_{\mathrm{R}}} - 1}. \tag{7}$$

The exponents ϕ_{L} and ϕ_{R} have to be identified from the TWC response to a step of the upstream lambda from lean to rich and vice versa.
Emission measurements are performed at the outlet of the catalyst, for which the lambda downstream catalyst λ_{ds} is pivotal. Due to the O_2 storage capability of the TWC, short lambda excursions away from the ideal stoichiometric point of $\lambda_{\mathrm{us}} = 1$ can be compensated and emissions can be kept low, with λ_{ds} remaining close to stoichiometry. This effect is reflected by the expression in Eq. (8) labelled as G_{Cat},

$$G_{\mathrm{Cat}}: \quad \lambda_{\mathrm{ds}} = \lambda_{\mathrm{us}} - (\lambda_{\mathrm{us}} - 1)\rho(\lambda_{\mathrm{us}}, ROL). \tag{8}$$

As long as $\rho \simeq 1$, which is the case for $ROL \approx 0.5$, λ_{ds} will remain close to 1 as well. If the ROL reaches either limit, ρ drops to 0 and $\lambda_{\mathrm{ds}} = \lambda_{\mathrm{us}}$. The target of an AFR controller that considers the oxygen storage of the TWC is therefore to keep the ROL in the region around 50 %, where the conversion processes of the TWC work most efficiently. A direct measurement of the current ROL is however not possible, which poses a challenging estimation problem from controls perspective.

2) Sensors: As indicated in Fig. 1, a typical AFR control setup contains lambda sensors upstream and downstream the TWC. For the upstream measurement usually a wideband (WB) lambda sensor is used in order to obtain quantitative information on the AFR. The engine tested for this work is equipped with a WB sensor of type Bosch LSU 4.9 TSP CV upstream the TWC.
The downstream lambda measurement can be set up with either a second wideband or a narrowband (NB) lambda sensor. The latter has the characteristic, that the sensor signal $U_{\lambda\mathrm{ds}}$ follows a switch-type curve, as illustrated in Fig. 2. For lean and rich operation, the sensor signal saturates, and it only provides qualitative information on whether the engine is running on a lean or rich AFR. In contrast, close to stoichiometry the signal gradient is very high, which results in a good sensor sensitivity. This is beneficial for control purposes, as it allows a more precise measurement of λ_{ds}, given a certain signal-to-noise ratio. For the test engine, both a WB and a NB sensor of type Bosch LSF 4.2 TSP downstream the catalyst is used.

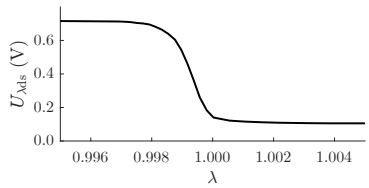


Fig. 2: Narrowband lambda sensor voltage signal $U_{\lambda \mathrm{ds}}$ [13].

According to Eq. (8), $\lambda_{\mathrm{ds}} \approx 1$ as long as the TWC oxygen storage is neither completely filled nor fully depleted. Hence, the sensor output will only change noticeably if the storage reaches one of the limits. Therefore, the sensor signal can be directly used as an approximation for the relative oxygen storage level [14],

$$ROL_{\mathrm{meas}} = \begin{cases} 1 & U_{\lambda \mathrm{ds}} \leq U_{\mathrm{low}}, \\ 1 - \frac{U_{\lambda \mathrm{ds}} - U_{\mathrm{low}}}{U_{\mathrm{high}}} & U_{\mathrm{low}} < U_{\lambda \mathrm{ds}} \leq U_{\mathrm{high}}, \\ 0 & U_{\lambda \mathrm{ds}} > U_{\mathrm{high}}. \end{cases} \quad (9)$$

Here U_{low} and U_{high} denote the saturation points of the sensor output voltage in Fig. 2. The actual O_2 storage level cannot be measured directly, therefore most AFR control schemes make use of this approximation.

D. Classical Control Approach

A simple method to form a closed loop λ control can be implemented with a cascaded structure of PID controllers. A fast inner loop with the controller K_{in} acts on the multiplier u_λ of the gas dosage and is designed to suppress a disturbance d in λ_{us}, cf. Fig. 3. The slow outer loop provides a set point $\lambda_{\mathrm{us,set}}$ for the inner loop, which also affects the gas dosage feedforward function. The corresponding controller K_{out} makes use of the feedback signal $U_{\lambda \mathrm{ds}}$ of the narrowband lambda sensor downstream the TWC and tries to reduce its deviation to a desired set point $U_{\lambda \mathrm{ds,set}}$.
This approach does not consider the storage dynamics of the TWC and can therefore only react on an emissions breakthrough rather than to predict and avoid it.

E. Robust H_∞ Control Approach

An improved catalyst conversion efficiency can be achieved if the oxygen storage capability of the TWC is taken into account in the controller design. For this purpose, the robust control approach discussed in [6] is considered in the following.
The basic structure of the control problem is illustrated in Fig. 3. In order to control the output $\Delta\lambda_{\mathrm{us}}$ of the engine fuel path G_λ from Eq. (3), a robust H_∞ controller K is designed that suppresses disturbances entering in d or d_{F}. A disturbance in the fuel dosage d_{F} is damped by the plant dynamics and is therefore less critical than a disturbance through channel d that can e.g. be caused by a cylinder misfiring. Due to that, the controller is optimised for disturbance rejection of input channel d.

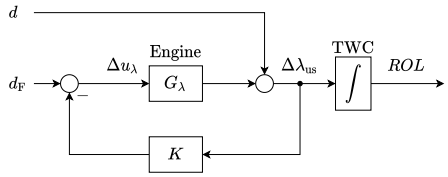


Fig. 3: Basic structure of control problem. The controller K measures the upstream catalyst $\Delta\lambda_{\mathrm{us}}$ and corrects the fuel dosage via a multiplicative factor $u_\lambda = \Delta u_\lambda + 1$. An estimate of the oxygen storage ROL in the TWC is included in the control design. Figure modified from ref. [6].

As discussed in Section II-C1, the ultimate goal of the control design is to reduce the influence of a disturbance d on the stored oxygen mass ROL of the TWC. Hence, the according transfer function $G_{\mathrm{TWC,d}}$ must be optimised for disturbance rejection,

$$G_{\mathrm{TWC,d}}(s) = \frac{ROL}{d} = \frac{1}{s\,(1 + G_\lambda(s)K(s))}. \quad (10)$$

An H_∞ controller is a promising approach for this purpose, due to its robustness against model uncertainties and guaranteed stability margins.

1) Padé Approximation: For the H_∞ control design, a linear plant model is required, which is not fulfilled by G_λ in Eq. (3) due to the contained delay term. This can be solved by approximating the delay with a Padé all-pass element.
The Padé approximation expresses the delay by a rational function in s of chosen order n_{apx} that best approximates the actual exponential behaviour. Equation (11) provides the instruction for computing the approximation.

$$\begin{aligned} e^{-s\delta} &= \frac{\sum_{k=0}^{n_{\mathrm{apx}}} a_k(-s\delta)^k}{\sum_{k=0}^{n_{\mathrm{apx}}} a_k(s\delta)^k} \\ a_k &= \frac{(2n_{\mathrm{apx}} - k)!\,n_{\mathrm{apx}}!}{k!(n_{\mathrm{apx}} - k)!(2n_{\mathrm{apx}})!} \end{aligned} \quad (11)$$

According to [6], the approximation order n_{apx} should be chosen depending on the delay time δ. For every $40\,\mathrm{ms}$ the order should be increased by 1, which can be also formulated as

$$n_{\mathrm{apx}}(\delta) = \left\lceil \frac{\delta}{40\,\mathrm{ms}} \right\rceil. \quad (12)$$

For typical delay times, the approximation order becomes very large. This is however not an issue, as the controller design includes an order reduction step at a later stage, as discussed in Section II-E5.

2) Augmented Plant: In a standard H_∞ control setup, the layout of a control circuit is interpreted in a very generalised fashion. A standard configuration is depicted in Fig. 4. The generalized or augmented plant P has two inputs, the generalized input w and the controlled input u. The latter is the output of the controller K and the former is typically a vector containing reference values, disturbance inputs, etc.
On the other hand, the plant outputs are the generalized error

(vector) z and the measured output y that serves as an input to the controller. Here, a single-input single-output system (SISO) is considered, therefore y and u are scalars. The aim of H_∞ control is to find a controller K that minimizes the influence of the disturbance input w on the error signal z.

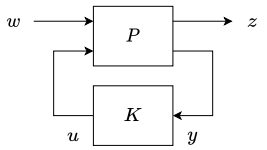


Fig. 4: Generalised control circuit in a typical H_∞ control setup with augmented plant P and controller K.

For the system structure in Fig. 3, the augmented plant can be constructed as shown in Fig. 5. It can be seen that the error vector z is formed with weighting functions $W_\mathrm{u}, W_\mathrm{y}$ and W_e. Those are used to shape the frequency response of the controller and are further discussed in Section II-E3.

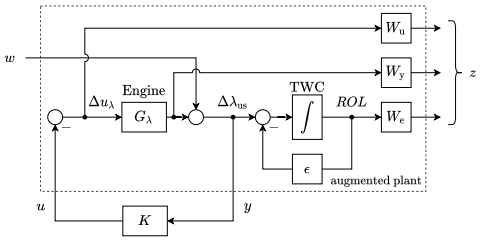


Fig. 5: Augmented plant for TWC control problem. Figure modified from ref. [6].

The ϵ term in the integrator representing the TWC is used for numerical stability of the optimization routine that tries to find the optimal controller K. Accordingly, a very small value of $\epsilon = 1 \times 10^{-4}$ is chosen to limit the low frequency gain of the integrator without having a negative impact on the controller design.

3) Weighting Functions: The behaviour of the synthesized H_∞ controller can be tuned by choosing appropriate error weighting functions. With W_u the controller effort KS can be penalised, W_e shapes the sensitivity function S and W_y the complementary sensitivity function T. S and T are defined as,

$$S = \frac{1}{1+GK}, \tag{13}$$

$$T = \frac{GK}{1+GK}. \tag{14}$$

For good reference tracking and disturbance rejection, the sensitivity S should be small at low frequencies. Noise attenuation and robustness can be optimized by making T small for high frequencies. The functions are connected through $S + T = 1$. The controller K determined by the H_∞ algorithm achieves a certain performance level γ. It is measured with the H_∞ norm of the closed loop transfer function P_cl formed between the augmented plant P and the controller K, cf. Fig. 4,

$$||P_\mathrm{cl}(s)||_\infty := \sup_{\omega\in\mathbb{R}} \sigma_\mathrm{max}\left\{P_\mathrm{cl}(j\omega)\right\} \overset{\mathrm{SISO}}{=} \sup_{\omega\in\mathbb{R}} |P_\mathrm{cl}(j\omega)| \leq \gamma, \tag{15}$$

with $\sigma_\mathrm{max}\{P_\mathrm{cl}(j\omega)\}$ being the largest singular value of the matrix $P_\mathrm{cl}(j\omega)$. In the case of a SISO system, this reduces to the maximum gain of the transfer function $P_\mathrm{cl}(j\omega)$.
With the achieved performance level, guaranteed margins for the TWC system can be formulated via the introduced weighting functions [6],

$$|G_\mathrm{TWC,d}(j\omega)| < \gamma \left|W_\mathrm{e}^{-1}(j\omega)\right|, \tag{16}$$

$$|T(j\omega)| < \gamma \left|W_\mathrm{y}^{-1}(j\omega)\right|. \tag{17}$$

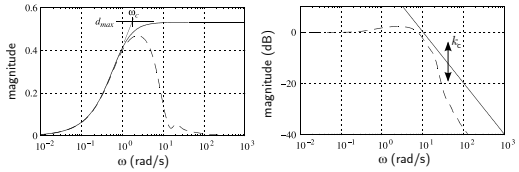
These relations highlight the importance of the weighting functions in H_∞ controller design. Their inverse serves as an upper bounds for the controlled system properties and can therefore be used for shaping the system response.
For the design in ref. [6], the weighting functions W_u, W_e and W_y are parametrised as follows,

$$W_\mathrm{u} = \epsilon, \tag{18}$$

$$W_\mathrm{e} = \frac{1}{d_\mathrm{max}} \frac{s+\omega_\mathrm{c}}{s+\epsilon}, \tag{19}$$

$$W_\mathrm{y} = \frac{s}{k_\mathrm{c}\omega_\mathrm{c}}. \tag{20}$$

By choosing the small gain coefficient $\epsilon = 1 \times 10^{-4}$ for the weighting function W_u, the controller effort limitation is effectively disabled. This choice is rather a measure to guarantee numerical stability of the solution process. The role of the parameters d_max, ω_c and k_c in shaping the system response is illustrated in Fig. 6.



(a) W_e^{-1} and $G_\mathrm{TWC,d}$ (dashed). (b) W_y^{-1} and T (dashed).

Fig. 6: Weighting functions of H_∞ controller design for TWC plant [6], [14].

According to ref. [6], the weighting function parameters can be related to the plant delay δ, cf. Eq. (3). This enables a stable performance and robustness across a wide parameter range. A value of $k_\mathrm{c} \approx 10$ is suggested and suitable parametrisation of d_max and ω_c is

$$d_\mathrm{max} = f_{d_\mathrm{max}}\delta, \tag{21}$$

$$\omega_\mathrm{c} = \frac{f_{\omega_\mathrm{c}}}{\delta}, \tag{22}$$

with $f_{d_\mathrm{max}} = 1.9$ and $f_{\omega_\mathrm{c}} = 0.22$.

4) H_∞ Controller Synthesis: The process of finding an optimal H_∞ controller for the TWC system is carried out in MATLAB. For this purpose, the command `hinfsyn` is used. The generalised system in Fig. 4 can be expressed by

$$\begin{pmatrix} z \\ y \end{pmatrix} = \begin{pmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{pmatrix} \begin{pmatrix} w \\ u \end{pmatrix}, \tag{23}$$

$$u = Ky. \tag{24}$$

Closing the loop via the controller K, the transfer function from w to z is found as

$$P_{\mathrm{cl}} = \frac{z}{w} = P_{11} + P_{12}K(I - P_{22}K)^{-1}P_{21}. \tag{25}$$

The algorithm numerically solves the optimization problem of finding a controller K that minimizes the H_∞ norm of P_{cl},

$$\gamma = \min_K ||P_{\mathrm{cl}}||_\infty , \tag{26}$$

which involves finding a solution for two Riccati equations.

5) Order Reduction: Due to the high system order resulting from the Padé approximation discussed in Section II-E1, an order reduction step of the determined H_∞ controller must be carried out to allow a practical implementation. A 4th order system is desired, as it suffices to reproduce the central aspects of the design [6].
The order reduction step is performed with a balanced model reduction routine in MATLAB, using the command `balred`.

6) Implementation Aspects: The fourth order controller obtained from the H_∞ design process results in 7 parameters $G_1 ... G_7$ that have to be gain scheduled over the engine load. Two integrators are contained in the controller, therefore the number of parameters of a general 4th order system reduces from 9 to 7.
Equation (27) describes the final structure of the controller and makes use of a maximally parallelized formulation, which allows a physical interpretation of the structure and the introduction of corrections [6].

$$K(s) = G_1 + G_2\frac{1}{s} + G_3\frac{1}{s^2} + G_4\frac{s(G_7 s + 1)}{G_5^2 s^2 + 2G_5 G_6 s + 1} \tag{27}$$

Fig. 7 shows the final layout of the controller. In the storage model segment, a modification is made to the TWC integrator. The controller is designed with the assumption of a normalised exhaust mass flow $\dot{m}_{\mathrm{exh}}$. In reality the actual mass flow has to be taken into account to reproduce the integrating behaviour of the oxygen storage correctly. The lambda deviation $\Delta\lambda_{\mathrm{us}}$ is therefore multiplied with the normalised oxygen mass fraction in the exhaust mass flow $0.23\dot{m}_{\mathrm{exh}}/C_{\mathrm{cat}}$. For being able to keep the normalised controller parameters, this correction has to be renormalised at the end of the storage model segment.
A second modification is the use of the downstream lambda sensor to obtain a measure for ROL according to the approximation in Eq. (9). The deviation of this measured value ROL_{meas} and the estimated stored oxygen level $\hat{ROL}$ is fed into a PI controller K_{corr} that creates a correction in the upstream λ_{us}, which effectively implements a corrector term for the estimated $\hat{ROL}$. This outer control loop has the effect of compensating a possible offset in the upstream lambda sensor or wrong assumptions on the storage capacity C_{cat}.

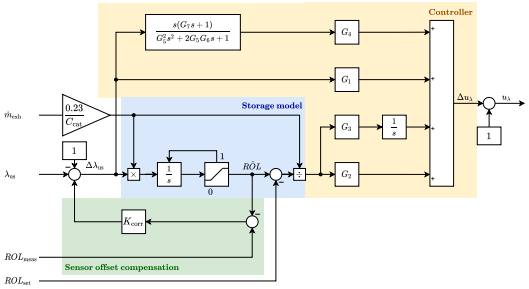


Fig. 7: Controller layout for the H_∞ robust control approach.

In a final step, the controller is discretised using the bilinear transform method. An automatic code generation routine transforms a MATLAB Simulink implementation of the design into code that can be run on a programmable logic controller (PLC) engine control system.

F. Adaptive Control

Over time, the TWC ceramics material might suffer from ageing and its storage and conversion capabilities can degrade. The control system should be able to handle such variations in the TWC system parameters and maintain a storage level in the optimum region of operation to ensure high conversion efficiency. A possible approach to solve this issue is the use of adaptive control techniques. Prior works by Shafai et al. [7] and Ammann et al. [8] are the basis of the work presented in the following.

1) TWC Parameter Estimation: The storage behaviour of the TWC system is primarily determined by the catalyst capacity C_{cat} and the storage and release rates K_{IP} and K_{IN}. The works in refs. [7], [8] suggest to only estimate the storage and release rates K_{IP} and K_{IN} and to keep C_{cat} fixed. In an aged TWC, the capacity C_{cat} would degrade, but the parameter estimation algorithm can create the same effect by increasing K_{IP} and K_{IN}, which is reported to be a more robust option.
From controls perspective, also the estimation of the true lambda upstream catalyst λ_{us} has a strong influence on how well the TWC model reproduces the actual behaviour. Furthermore, it is known that the wideband lambda sensor tends to measure a biased λ_{us} compared to the true value $\tilde{\lambda}_{\mathrm{us}}$ [7],

$$\tilde{\lambda}_{\mathrm{us}} = \lambda_{\mathrm{us}} + K_\lambda, \tag{28}$$

where K_λ denotes the signal bias.
In combination, the adaptive control algorithm is used to estimate three system parameters,

$$\hat{\theta} = (\hat{K}_{\mathrm{IP}} \ \hat{K}_{\mathrm{IN}} \ \hat{K}_\lambda)^{\mathrm{T}}, \tag{29}$$

with $\hat{\theta}$ being the parameter vector used in the optimization routine in Section II-F2 and estimated parameters are marked with a hat symbol.

Fig. 8 shows the necessary modification in the TWC storage estimator structure. The estimated signal bias $\hat{K}_\lambda$ is removed from the measured signal. Depending on whether the resulting $\Delta\tilde{\lambda}_{\mathrm{us}}$ is in the lean (≥ 0) or the rich (< 0) region, the according rate coefficient $\hat{K}_{\mathrm{IP}}$ or $\hat{K}_{\mathrm{IN}}$ is used as integrator gain.

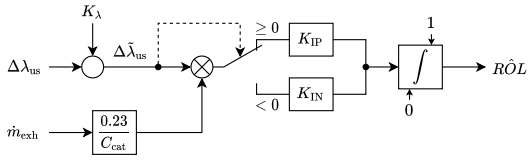


Fig. 8: TWC storage model with parameter estimation. The system parameters to be adapted are K_{IP}, K_{IN} and K_λ. Figure modified from ref. [7].

The adaptation routine is carried out in a discrete event fashion. By using the downstream lambda sensor the occasions where a breakthrough in the TWC happens are detected. Such events are identified by introducing voltage limits for the narrowband lambda sensor that define a lean or rich breakthrough. Comparing with Fig. 2, the limits can be chosen as $U_{\lambda\mathrm{ds}} \leq U_{\mathrm{low}} = 100\,\mathrm{mV}$ and $U_{\lambda\mathrm{ds}} \geq U_{\mathrm{high}} = 750\,\mathrm{mV}$. When a limit is exceeded, a parameter estimation update is triggered at the according timing point t_k. In case of a lean breakthrough $ROL(k) = 1$ and for a rich breakthrough $ROL(k) = 0$.
For the parameter estimation, the TWC storage integrator is modelled in the time periods between two trigger events. The estimate $R\hat{O}L$ is determined from

$$\begin{aligned} R\hat{O}L(k) = ROL(k-1) + (I_{\mathrm{P}}(k) + I_{\mathrm{C1}}(k)\hat{K}_\lambda)\hat{K}_{\mathrm{IP}} + \\ + (I_{\mathrm{N}}(k) + I_{\mathrm{C2}}(k)\hat{K}_\lambda)\hat{K}_{\mathrm{IN}}, \end{aligned} \tag{30}$$

with the following integrals having to be calculated online,

$$I_{\mathrm{P}}(k) = \int_{t_{k-1}}^{t_k} \left(\frac{0.23}{C_{\mathrm{cat}}}\dot{m}_{\mathrm{exh}}(t)\Delta\lambda_{\mathrm{us}}(t)\right)\Big|_{\Delta\lambda_{\mathrm{us}}\geq 0} \mathrm{d}t, \tag{31}$$

$$I_{\mathrm{N}}(k) = \int_{t_{k-1}}^{t_k} \left(\frac{0.23}{C_{\mathrm{cat}}}\dot{m}_{\mathrm{exh}}(t)\Delta\lambda_{\mathrm{us}}(t)\right)\Big|_{\Delta\lambda_{\mathrm{us}}< 0} \mathrm{d}t, \tag{32}$$

$$I_{\mathrm{C1}}(k) = \int_{t_{k-1}}^{t_k} \left(\frac{0.23}{C_{\mathrm{cat}}}\dot{m}_{\mathrm{exh}}(t)\right)\Big|_{\Delta\lambda_{\mathrm{us}}\geq 0} \mathrm{d}t, \tag{33}$$

$$I_{\mathrm{C2}}(k) = \int_{t_{k-1}}^{t_k} \left(\frac{0.23}{C_{\mathrm{cat}}}\dot{m}_{\mathrm{exh}}(t)\right)\Big|_{\Delta\lambda_{\mathrm{us}}< 0} \mathrm{d}t. \tag{34}$$

Due to the occurring product terms between $\hat{K}_{\mathrm{IP}}$, $\hat{K}_{\mathrm{IN}}$ and $\hat{K}_\lambda$, the parameter estimation problem is non-linear.

2) Recursive Gauss-Newton Algorithm: A possible solution procedure to the non-linear parameter estimation problem is the Recursive Gauss-Newton algorithm. Between two trigger events, the change in ROL, $y(k) = ROL(k) - ROL(k-1)$, can only take the following values,

$$y(k) = \begin{cases} 1 & U_{\mathrm{low}} \rightarrow U_{\mathrm{high}}, \\ -1 & U_{\mathrm{high}} \rightarrow U_{\mathrm{low}}, \\ 0 & U_{\mathrm{low}} \rightarrow U_{\mathrm{low}} \text{ or } U_{\mathrm{high}} \rightarrow U_{\mathrm{high}}. \end{cases} \tag{35}$$

The non-linear parameter identification problem can be stated as

$$y(k) = \phi^{\mathrm{T}}(k) f(\hat{\theta}) + e(k), \tag{36}$$

where $e(k)$ corresponds to a white noise term whose variance is estimated online and the remaining terms are defined as follows,

$$\phi^{\mathrm{T}}(k) = (I_{\mathrm{P}}(k)\ I_{\mathrm{N}}(k)\ I_{\mathrm{C1}}(k)\ I_{\mathrm{C2}}(k)), \tag{37}$$

$$f(\hat{\theta}) = (\hat{\theta}_1\ \ \hat{\theta}_2\ \ \hat{\theta}_1\hat{\theta}_3\ \ \hat{\theta}_2\hat{\theta}_3)^{\mathrm{T}}. \tag{38}$$

The goal of the recursive Gauss-Newton algorithm is to numerically search for a zero crossing of the error term

$$\epsilon(k, \hat{\theta}(k-1)) = y(k) - \phi^{\mathrm{T}}(k) f(\hat{\theta}(k-1)). \tag{39}$$

An implementation of the algorithm is presented in Equations (40) to (46).

$$\psi(k) = \begin{pmatrix} I_{\mathrm{P}}(k) + I_{\mathrm{C1}}(k)\hat{\theta}_3(k-1) \\ I_{\mathrm{N}}(k) + I_{\mathrm{C2}}(k)\hat{\theta}_3(k-1) \\ I_{\mathrm{C1}}(k)\hat{\theta}_1(k-1) + I_{\mathrm{C2}}(k)\hat{\theta}_2(k-1) \end{pmatrix} \tag{40}$$

$$\kappa(k) = \frac{\kappa(k-1)}{\beta + \kappa(k-1)} \tag{41}$$

$$Q(k) = Q(k-1) + \kappa(k)\left[\epsilon^2(k, \hat{\theta}(k-1)) - Q(k-1)\right] \tag{42}$$

$$S(k) = \beta Q(k) + \psi^{\mathrm{T}}(k) P(k-1) \psi(k) \tag{43}$$

$$L(k) = P(k-1)\psi(k) S^{-1}(k) \tag{44}$$

$$\hat{\theta}(k) = \hat{\theta}(k-1) + L(k)\epsilon(k, \hat{\theta}(k-1)) \tag{45}$$

$$P(k) = \frac{1}{\beta}\left[P(k-1) - L(k)S(k)L^{\mathrm{T}}(k)\right] \tag{46}$$

As a starting point, the algorithm needs initial values for the parameter estimation vector $\hat{\theta}(0)$ as well as the covariance matrix $P(0)$. The latter is typically set to a diagonal matrix with large entries, which indicates low confidence in the initial parameter guess. Finally, β has the role of a forgetting factor. If it is set to 1, the recursive algorithm takes all historic values into account, whereas for a typical value of 0.99 a slow forgetting rate can be implemented.

III. Results

Due to the conflict of interests and as protection of proprietary information, all measurements presented are normalised to set points or variables of withheld magnitude, labelled by indices set or ref, respectively.

A. AFR Control

The disturbance rejection properties of the discussed control approaches in Sections II-D and II-E for AFR control are tested in steady state by introducing a gas dosage perturbation via the multiplier u_λ, i.e. channel d_{F} in Fig. 3. Moreover, the transient performance is tested by performing sweeps of the engine load P_{norm}. A comparison of the disturbance rejection results for the classical PID approach and the H_∞ controller is illustrated in Fig. 9. Results on transient performance are shown in Fig. 10.

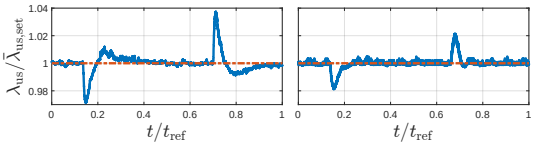


Fig. 9: Disturbance rejection performance. Left: H_∞, right: cascade PID. The dashed line marks the lambda set point $\bar{\lambda}_{\mathrm{us,set}}$. Measurements are aligned with respect to the positive fuel multiplier u_λ disturbance step at $t_{\mathrm{ref}}/16$.

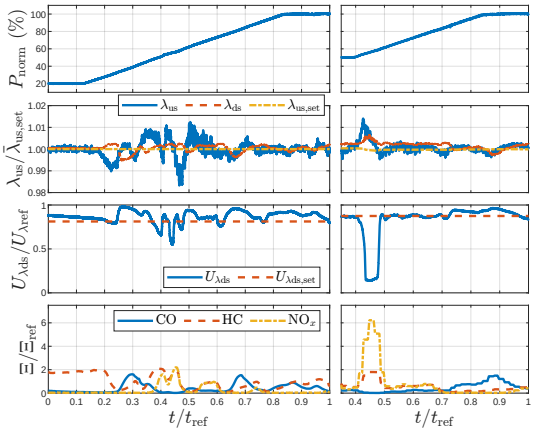


Fig. 10: Transient performance during load ramp. Left: H_∞, right: cascade PID. 1st plot: load ramp of power P_{norm}, 2nd plot: Lambda engine out λ_{us} with set point $\lambda_{\mathrm{us,set}}$ (dash-dotted) and catalyst out lambda λ_{ds} (dashed), 3rd plot: NB lambda sensor voltage $U_{\lambda\mathrm{ds}}$ with set point $U_{\lambda\mathrm{ds,set}}$ (dashed), 4th plot: mass concentration emissions Ξ for CO, HC (dashed) and NO_x (dash-dotted).

It is apparent, that the H_∞ and PID controllers react differently to a disturbance. In the H_∞ design, a deflection of lambda in the opposite direction of the disturbance ensures, that the oxygen storage deviations are balanced, which is beneficial for the TWC efficiency. On the other hand, the cascade PID controller achieves lower overshoot and faster settling time.
The transient test illustrates, that the lambda variation is comparable in both cases. Also, the damping effect on λ_{ds} due to the oxygen storage effect of the catalyst can be observed. For the H_∞ controller the HC emissions are initially relatively high, whereas a lean NO_x breakthrough can be observed for the PID design at $t = 0.4t_{\mathrm{ref}}$, accompanied by a dip in the NB sensor voltage $U_{\lambda\mathrm{ds}}$. At steady state, both designs achieve comparable emission levels, hence neither of the two has a decisive performance advantage.

B. ROL Estimation

Recordings from an engine trial run are used to test the oxygen storage model used in the H_∞ and adaptive control approaches. The behaviour of the $R\hat{O}L$ estimate compared to the NB sensor approximated measurement ROL_{meas} from Eq. (9) is illustrated in Fig. 11.

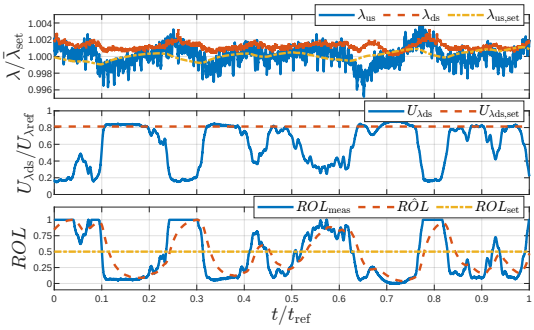


Fig. 11: TWC oxygen storage observer. 1st plot: Lambda upstream λ_{us} with inner loop set point $\lambda_{\mathrm{us,set}}$ (dash-dotted) and downstream catalyst λ_{ds} (dashed). 2nd plot: NB lambda sensor voltage $U_{\lambda\mathrm{ds}}$ with set point $U_{\lambda\mathrm{ds,set}}$ (dashed), 3rd plot: TWC oxygen storage level measurement ROL_{meas}, observer estimate $R\hat{O}L$ (dashed) and set point ROL_{set} (dash-dotted).

Although lambda is held at a stable operating point $\lambda_{\mathrm{us,set}}$, frequent toggling of the NB sensor signal $U_{\lambda\mathrm{ds}}$ is observed. This indicates, that small excursions in λ_{us} can lead to saturation and depletion of the TWC oxygen storage. Thus, the storage capacity C_{cat} is expected to be relatively small.
The feedback correction to the observer $R\hat{O}L$ through ROL_{meas} is tuned such, that the measurement is trusted more than the model. Accordingly, $R\hat{O}L$ closely follows the measurement ROL_{meas}. As the H_∞ controller generates a gas dosage correction based on the $R\hat{O}L$ estimate, a tuning that weights the NB measurement stronger leads to a more aggressive control behaviour. The illustrated setting is considered as a good compromise between controller aggressiveness and noise suppression.

C. Adaptive Control

In order to test the TWC parameter estimation in a simulation, a sinusoidal excitation is added to the gas dosage to enforce a cycling of the ROL through the saturation limits.
The results presented in Fig. 12 are obtained for a catalyst model with $K_{\mathrm{IP}} = 6.0$, $K_{\mathrm{IN}} = 4.0$ and a wideband lambda sensor bias of $K_\lambda = -3 \times 10^{-4}$. The forgetting factor $\beta = 0.95$, the parameter vector is initialised with $\theta(0) = (1\ 1\ 0)^{\mathrm{T}}$ and the covariance matrix with $P(0) = 0.3I$.
Fig. 12 illustrates, that a poorly conditioned adaptive control scheme requires many excitations to converge towards an appropriate parameter set. For $\hat{K}_{\mathrm{IP}}$ and $\hat{K}_\lambda$ it is observed, that the estimation error initially increases. When the algorithm has converged to a stable solution, the parameters are estimated sufficiently well, but small deviations to the true parameter values remain. This effect is more pronounced for the storage and release rates $\hat{K}_{\mathrm{IP}}$ and $\hat{K}_{\mathrm{IN}}$ than for the sensor bias $\hat{K}_\lambda$. The residual deviation is attributed to the simulation model

for λ_{ds} in Eq. (8), which is used to derive the narrowband sensor signal. An influence of the oxygen storage and release probability functions f_{L} and f_{R} from Eqs. (6) and (7) is not considered in the $\hat{ROL}$ estimator model, which results in a systematic estimation error for K_{IP} and K_{IN}. As the ROL behaviour can be reproduced well, this approximation error can be tolerated.

It is observed that the non-linearity in the parameters of the model can lead to diverging estimations. Therefore, care must be taken in properly initialising the system parameters and a large value for β, i.e. a low forgetting rate, is recommended to avoid fast parameter drifts.

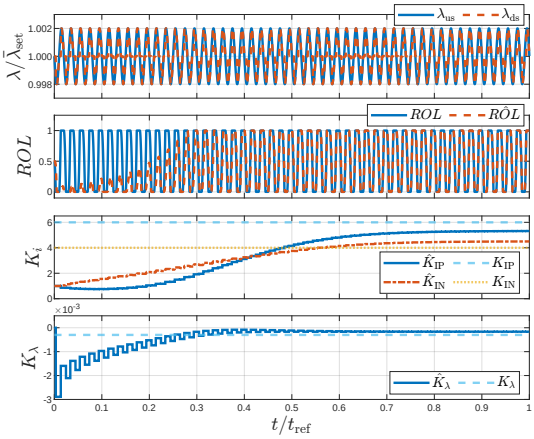


Fig. 12: Adaptive control TWC parameter estimation. 1st plot: Lambda upstream λ_{us} and downstream catalyst λ_{ds} (dashed). 2nd plot: Actual ROL of the TWC and observer $\hat{ROL}$ (dashed). 3rd plot: Estimated value for storage $\hat{K}_{\mathrm{IP}}$ and release rate $\hat{K}_{\mathrm{IN}}$ (dash-dotted) with true values K_{IP} (dashed) and K_{IN} (dotted). 4th plot: Estimated value for upstream lambda sensor bias $\hat{K}_\lambda$ and true value K_λ (dashed).

IV. Conclusion

The presented work demonstrates, that H_∞ robust control is a promising concept for efficient AFR control that can consider the oxygen storage capabilities of the TWC. In experiments, a too small storage capacity C_{cat} prevents that the ROL can be stabilised at a desired value. Hence, the potential for emission reduction compared to a classical approach cannot be observed.

Furthermore, TWC parameter estimation is demonstrated. This enables an adaptive control setting that keeps track of system parameters and therefore reflects and incorporates changes to component behaviour e.g. due to ageing effects.

Acknowledgment

The authors would like to thank the company INNIO Jenbacher GmbH & Co OG for providing the facility and prototype engine to carry out the tests on engine and aftertreatment system. Special thanks go to the MCI Management Center Innsbruck for enabling a fruitful cooperation and creating a motivating and inspiring place to work on a graduate study programme.

References

[1] *Vierundvierzigste Verordnung zur Durchführung des Bundes-Immissionsschutzgesetzes (Verordnung über mittelgroße Feuerungs-Gasturbinen- und Verbrennungsmotoranlagen - 44.BImSchV)*, Bundesministerium für Umwelt, Naturschutz und nukleare Sicherheit - BMU Deutschland Std., (BGBl. I S. 804).

[2] R. W. Schallock, K. R. Muske, and J. C. P. Jones, "Model predictive functional control for an automotive three-way catalyst," *SAE International Journal of Fuels and Lubricants*, vol. 2, no. 1, pp. 242–249, 2009.

[3] S. Trimboli, S. Di Cairano *et al.*, "Model predictive control for automotive time-delay processes: An application to air-to-fuel ratio control," *IFAC Proceedings Volumes*, vol. 42, no. 14, pp. 90–95, 2009, 8th IFAC Workshop on Time-Delay Systems.

[4] E. Shafai, C. A. Roduner, and H. Geering, "On-line identification of time delay in the fuel path of an si engine," SAE Technical Paper, Tech. Rep., 1997.

[5] G. A. Ingram and G. Surnilla, "On-line oxygen storage capacity estimation of a catalyst," SAE Technical Paper, Tech. Rep., 2003.

[6] C. Roduner, C. Onder, and H. Geering, "Automated design of an air/fuel controller for an SI engine considering the three-way catalytic converter in the H∞ approach," in *Proceedings of the fifth IEEE Mediterranean Conference on Control and Systems, Paphos, Cyprus*, 1997.

[7] E. Shafai, C. Roduner, and H. P. Geering, "Indirect adaptive control of a three-way catalyst," in *International Congress & Exposition*. SAE International, feb 1996.

[8] M. Ammann, H. Geering *et al.*, "Adaptive control of a three-way catalytic converter," in *Proceedings of the 2000 American Control Conference. ACC (IEEE Cat. No.00CH36334)*, vol. 3, 2000, pp. 1561–1566 vol.3.

[9] J. Rink, N. Meister *et al.*, "Oxygen storage in three-way-catalysts is an equilibrium controlled process: Experimental investigation of the redox thermodynamics," *Applied Catalysis B: Environmental*, vol. 206, pp. 104–114, 2017.

[10] M. I. Soumelidis, R. K. Stobart, and R. A. Jackson, "A nonlinear dynamic model for three-way catalyst control and diagnosis," *SAE Transactions*, vol. 113, pp. 764–775, 2004.

[11] E. P. Brandt, Y. Wang, and J. W. Grizzle, "A simplified three-way catalyst model for use in on-board SI engine control and diagnostics," in *Proceedings of the ASME Dynamic Systems and Control*, 1997, pp. 653–659.

[12] E. Brandt, Y. Wang, and J. Grizzle, "Dynamic modeling of a three-way catalyst for SI engine exhaust emission control," *IEEE Transactions on Control Systems Technology*, vol. 8, no. 5, pp. 767–776, 2000.

[13] Robert Bosch GmbH, *Oxygen sensor LSF 4.2 TSP*, Technical Customer Documentation.

[14] L. Guzzella and C. H. Onder, *Introduction to Modeling and Control of Internal Combustion Engine Systems*, 2nd ed. Berlin: Springer, Berlin, Heidelberg, 2010.

Michael Schmid is master student at the Mechatronics department of the MCI Management Center Innsbruck. Before joining the part time program of "Mechatronics & Smart Technologies", he studied physics at the University of Innsbruck with a major in quantum physics. Michael is working as a PLC programmer and controls engineer at the Tyrolean company INNIO Jenbacher, which is specialised on building gas engine combined heat and power (CHP) plants.

Simplified Ultimate Load Calculation of Real Steel-Structures in Earth-Moving Machinery

Nico Schuchter, Claus Lamprecht (supervisor), and Franz-Josef Falkner (supervisor)

***Abstract*—Due to time and cost savings mainly linear finite element analysis (FEA) are performed to determine the structural strength of components used in earth-moving machinery. This has the disadvantage that no statement about the actual structural load limit can be made, since non-linear plastic material behaviour is not considered. The latter is explicitly allowed in some standards applicable to the machine type under consideration. The reason for this is to provide manufacturers' with a possibility to be able to meet the high safety requirements given in the standards, which guarantee a safety against structural failure.**

This paper investigates and compares two approaches on how to determine plastic material reserves by taking non-linear material behaviour into account. The more accurate but also more time consuming approach is to perform non-linear FEA. The second, more conservative approach is the methodology given in the FKM guideline. A comparison of both gives an indication as to whether non-linear FE-analyses are actually necessary or if the linear approach is sufficient enough.

In the end it can be seen, that the actual load limits of the investigated components can only be determined by non-linear FEA. However, it is possible to develop a methodology to establish pre-defined strength requirements based on the results of the non-linear FEA. These requirements can then be translated to similar machine components and problems. This gives engineers the possibility to estimate plastic material reserves and guarantee compliance with applicable standards. All without having to perform explicit non-linear FEA for each component to be investigated in the future. Furthermore, it was possible to ensure that all the components examined met the standards and structural failure is avoided. More generally spoken, it can be said that depending on the component and load case, enormous plastic reserves are available. Hence, in some cases the additional effort of a non-linear FEA is worthwhile, as unexpected material reserves are discovered.

***Index Terms*—FEM, earth-moving machinery, linear & non-linear material behaviour, linear & non-linear FEA, Neuber rule, FKM guideline, plastic support effect.**

I. Introduction

STRUCTURAL mechanical strength verifications are usually based on applying the finite element method (FEM) and the accompanying FEA. Both play an essential role in the technical development work of earth-moving machinery. The term earth-moving machinery describes self-propelled or towed machines on wheels, crawlers or tracks with an operating weight of at least $4500\,\mathrm{kg}$ and mechanical fixtures to mount equipment. Examples include crawler tractors, telescopic handlers and special-purpose machines like pipe layers. Those types of machines are used in various industries, e.g. agriculture, construction, oil & gas industry or open-cast mining, and must meet high safety standards [1].

As a part of a review to check compliance with standards applicable to the considered machine type, certain points arise that this paper deals with:

- Necessity of non-linear FEM-analyses.
- Utilisation of plastic material behaviour.
- Detail accuracy of the models used for FEM/FEA.
- Additional work and time expenditure.

International standards like ISO-15818 [2] or ISO-10532 [3] specify certain safety factors on specific machine components against structural mechanical failure. These safety factors can be as high as four times of the initial load applied, before mechanical failure is allowed. The focus is set on machine components like lifting & tying-down attachment points or retrieval hooks and lugs which are used in shipment, loading or recovery operations, see Figure 1.

(a) Front lashing point. (b) Rear lashing and retrieval point.

Fig. 1. Example of an eyehook serving as a front lashing point (a) and as a rear lashing & retrieval point (b) of a crawler tractor, both taken from [4].

Generally, and due to time and cost savings mainly linear FEA are performed to verify machine components compliance with the applicable standards. This has the disadvantage, that there is no possibility to determine the component's actual load limit, where non-linear plastic behaviour comes into account. The applicable standards allow this kind of material behaviour and even presuppose it in order to be able to attain the to-be-achieved safety factors.

These plastic material reserves are not to be neglected, since the resulting equivalent stresses σ_V are allowed to exceed the material's yield strength R_e significantly. This behaviour is described with the help of the plastic support number n_{pl}. It gives an indication on how far a component can be loaded beyond the elastic load limit L_{el}, where the stress state $\sigma_V = R_e$ is reached for the first time.

In order to determine these material reserves non-linear FEA have to be conducted. The aim of this research is to investigate non-linear material behaviour of components that are loaded beyond their material's R_e. The general objective is to determine the non-linear load limits for multiple machine components and their compliance with applicable standards. In addition, the influence of different weld modelling concepts is also investigated.

II. Computational Models

The preparation of the computational models for the linear and non-linear FEA is essential for the quality of the results. The modelling and abstraction of a real problem always requires certain assumptions and simplifications, in order to make the problem accessible to a solution with existing established methods like FEM. This chapter describes the preparation of the 3D volume geometry designed in a computer-aided design (CAD) program to the volume model used for FEA, exemplarily shown with one specific component.

A. Weld Modelling

The modelling of welds is based on the nominal stress, structural stress and notch stress concept mentioned in [5]. The concepts mainly differ in the modelling effort and the accuracy of the modelled weld. In the course of this work only fillet welds, see Figure 2, are considered.

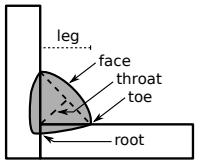
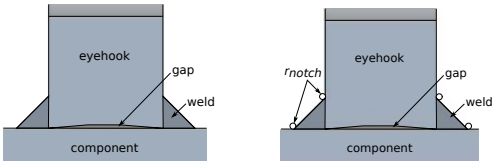


Fig. 2. There are five pieces to a fillet weld known as the root, toe, face, leg and throat, usually given in technical drawings as weld thickness a [6].

The nominal stress concept does not consider the effects of a weld, hence there is no actual visual modelling of the weld. The components that are actually welded together are seen as one. This is the standard modelling approach for structural strength assessments and verifications. The structural stress concept considers welds and their effects. But the weld is modelled in a simplified manner, in form of a right-angled triangular cross-section along the contact edges of the two components. Additionally, a gap is added at where the contact surface would be. This is necessary to allow the force flow to pass over the modelled weld, see Figure 3a. The modelling process of the notch stress concept is similar to the structural stress concept, see Figure 3b. The difference is that a fictional notch radius of $r_{notch} = 1\,\mathrm{mm}$ is added to the weld toes and other weld edges. The exact modelling of the weld root is only necessary if the root stresses are of interest or if a failure at the weld root is assumed. In the course of this work an exact modelling of the weld root is dispensed.



(a) Structural stress concept. (b) Notch stress concept.

Fig. 3. Weld modulation according to the structural stress concept (a) and the notch stress concept(b).

B. Force Determination

The forces to be applied are determined with the help of the standards and the load cases described therein. In general, the determination and calculation of the applied load respectively of the force components is based on trigonometry and vector calculation. Additionally, the standards specify two factors that have to be considered in the calculations. The first is the so-called proof factor $S_{pf} = 1.50$, which guarantees safety at the material's yield strength R_e. The second is the so-called breaking factor $S_{bf} = 4.00$ for safety against mechanical failure. Both factors are used to calculate the proof force F_{proof} and the breaking force F_{break}. In Table I a summary of all proof and breaking factors that have to be considered is given.

TABLE I
Overview of proof and breaking factors given in [2], [3] and [7].

Operation	Proof factor S_{pf}	Breaking factor S_{bf}
Lashing acc. to [2]	1.25	2.00
Lifting acc. to [2]	1.50	4.00
Retrieval acc. to [3]	1.50	-
Retrieval acc. to [7]	-	1.50

For all components and structures investigated, the load is applied to eyehooks intended for lashing, lifting, recovering or securing, see Figure 1. The load is transmitted to the eyehook via a chain, rope, pin or a shackle-pin connection.

C. Investigated Components & Structures

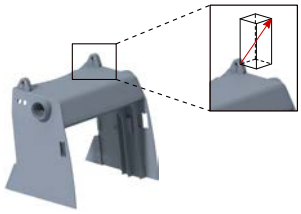
Various components and structures of crawler tractors and telescopic handlers are investigated, see Table II below. For the sake of brevity only one component, marked in Table II, is described in more detail, since it is also used in chapter IV for a detailed evaluation.

TABLE II
Overview and location of the investigated components.

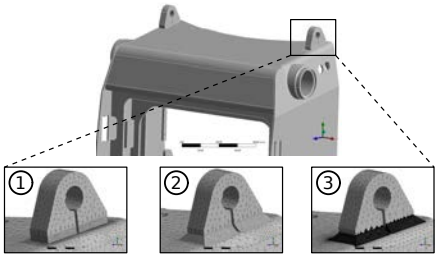
Component	Machine	Location
Counterweight	PR726	Rear attachment mounts
Counterweight	PR746	Rear attachment mounts
Counterweight	PR766	Rear attachment mounts
Fall protection	PR776	Welded onto cabin panels
Lifting point	PR746	Support frame's upper front area
Cabin lifting point	PR736	Bolted onto cabin roof
Lifting point	T55-7	Support frame's upper rear area
Lashing point	T55-7	Support frame's rear quarter panel

The support frame of the crawler tractor PR746 was subjected to a real structural test to investigate the frame's behaviour during loading operations. More specifically, the lifting point at the upper front area, see Figure 4, was examined. The behaviour of the eyehook as well as of the substructure is particularly interesting, especially when applying high loads like defined in [2]. In addition to this real test, a simulative approach by applying linear and non-linear FEA is also conducted. The aim is to create a comparison between the real structural test and simulative/numerical approach.To keep the simulation model simple a submodel of the supports frame's upper front area is designed, see Figure 4a. The load case under investigation is an oblique upwards pull, see the detail

in Figure 4a. Since the eyehook is welded onto the upper steel-plate, the concepts for modelling of welds must be observed and applied, see Figure 4b below.



(a) Simplified CAD submodel of the support frame's front area with visualization of the investigated load case.



(b) Ansys model of the support frame with applied weld modelling concepts: nominal ①, structural ②, and notch stress concept ③.

Fig. 4. Simplified submodel of the support frame's front area used for FEA.

III. Applied Methodology

This chapter describes the way to the actual solution of the FEM problems. It shows the general approach, modelling in Ansys and presents selected results of the aforementioned machine component.

Fig. 5. Ansys *Workbench* project for one component and load case.

The project overview from Ansys *Workbench* in Figure 5 above shows the typical structure for a FEA of one component and one specific load case. At the top left one can see the imported CAD geometry, which represents the starting point. The project is then set up with the help of the analysis systems of the type *Static Structural*. The top row in the rectangle with a solid line contains the linear finite element (FE)-simulations. Below and inside the dashed line are the non-linear FE-simulations.

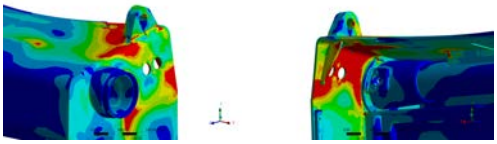
A. Linear & Non-linear FEA

Linear FEA forms the basis for investigating the static strength of all machine components. It provides initial information about the effect of a load acting on the component and serves to determine critical areas. Reaching or even exceeding the materials R_e should be avoided. This guarantees that the component or structure will withstand the applied loads. However, depending on where the critical areas are and the degree of simplification of the modelling, stresses above R_e can also be permitted without any impairment or even failure being assumed. The disadvantage is that only effects in the range of linear material behaviour can be considered. Therefore, no statement about non-linear effects such as plastic strain and deformation can be made. Standards such as [2] or [3] allow plastic material behaviour. Hence, it is difficult or even impossible to make a reliable statement.

Non-linear FEA represent an extended investigation of, in this case, statically loaded components and structures. Compared to the static linear FEA, which is in principle time-independent, the non-linear FEA takes into account time-dependent effects such as creep. However, since no inertia effects are taken into account, it can still be counted among the static analyses. Nonetheless, it is decisive that the character of the non-linearity should be captured. For this purpose, the load must be applied in steps or increments [8]. The advantages of a non-linear FEA are, as also mentioned in [9] and [10], that the plastic behaviour is taken into account. This is of great importance, for example, to fulfil standards like [2] or [3]. The biggest disadvantages, apart from the additional time-exposure required, are the difficult interpretation of the results and errors that can arise due to the simplified modelling and wrong settings.

B. Identifying Critical Areas

The first step of the post-processing procedure is to identify the critical areas. This is necessary in order to specify where component failure could occur or at least must be assumed. The maximum equivalent stress σ_V in MPa, the deformation d in mm and the plastic strain ε_{pl} in $\%$ are used as decisive variables. In general, the higher the level of model detail,



(a) (b)

Fig. 6. Stress distribution at the support frame's upper front area after applying the diagonal upwards pull; shown is the model with applied nominal stress concept.

the higher the resulting σ_V. Additionally, the meshing also has an effect and can lead to very high stresses at corners and edges, which do not occur in reality. Especially for the nominal stress concept, it must be taken into account that the stresses at the $90°$ transition are due to singularity and should

not be used for evaluations, see Figure 6. It turns out that the weld and the eyehook are the areas that experience the highest stresses. This holds true for all investigated components. In most cases, the weld and not the eyehook is the component where a possible failure must be assumed. Therefore, these two parts in particular are taken into account and evaluated in the following.

C. Determination of the Plastic Support Number

In order to be able to evaluate the critical areas, the static strength of the weld and the eyehook is assessed. This gives the possibility to make a statement about the maximum load capacity of the examined component. Conclusions can then be drawn from this about the requirements to be met in the standards. This component-specific result can then also be used to draw conclusions and make statements about similar components from other machine model series. However, this is the content of chapter IV.

The static strength of the weld and the eyehook are assessed with the aid of the plastic support number n_{pl}. The latter is determined by considering the elongation criterion K'_p and the plastic form factor K_p, see formula (1) below. All three weld modelling concepts as well as both critical areas, weld and eyehook, are evaluated individually. The evaluation is based on the results of the linear and non-linear FEA.

$$n_{pl} = \min(K'_p; K_p) = \min\left(\sqrt{\frac{E\varepsilon_{tol}}{R_e}}; \frac{L_{pl}}{L_{el}}\right) \tag{1}$$

1) Elongation Criterion: This determination of K'_p is based on the linear FEA and strictly adheres to [9]. This represents the more conservative approach, which gives little to no information about the components actual maximum load limit. It is based on [11] and assumes linear material behaviour. The calculation process of K'_p is independent of the applied weld modulation concept. In the first step, σ_V and σ_H are determined from the linear FEA. The former is defined as

$$\sigma_h = \frac{\sigma_x + \sigma_y + \sigma_z}{3} \tag{2}$$

according to [8]. The multiaxiality h can be determined by

$$h = \frac{\sigma_H}{\sigma_V} = \frac{\sigma_x + \sigma_y + \sigma_z}{3\sigma_V}. \tag{3}$$

With the help of the definition $\varepsilon_{ref} = A_{\%}$ and h the tolerable plastic strain ε_{tol} at the critical area can be determined, see formula (4) below.

$$\varepsilon_{tol} = \begin{cases} \varepsilon_{ref} & \text{for} \quad h \leq 1/3 \\ \varepsilon_0 + 0.3\left(\frac{\varepsilon_{ref} - \varepsilon_0}{0.3}\right)^{3h} & \text{for} \quad h > 1/3 \end{cases} \tag{4}$$

Table III shows all specified strains in dependence on the weld modelling concept. The structural stress concept is an exception, here $\varepsilon_{tol} = 5\,\%$ always applies. The tolerable equivalent stress according to the elongation criterion is defined as $\sigma_{tol} = K'_p R_e$.

TABLE III
SPECIFIED STRAINS FOR EVALUATION OF K'_p ACCORDING TO [9].

	Nominal Stress	Structural Stress	Notch Stress
ε_0	$5\,\%$	$5\,\%$	$5\,\%$
ε_{ref}	$A_{\%}$	$A_{\%}$	$17\,\%$

2) Plastic Form Factor: The determination K_p is based on the non-linear FEA and takes [10] into account. This leaves the engineers a certain amount of leeway in which own limits can be specified. K_p is determined by definition from the ratio of the plastic load limit L_{pl} and the elastic load limit L_{el}. Due to the unitlessness it does not matter whether an actual load L in N or a load factor (LF) is used. Thus, K_p can alternatively be defined as

$$K_p = \frac{LF_{pl}}{LF_{el}} = \frac{L_{pl}}{L_{el}}. \tag{5}$$

The elastic load factor LF_{el} can be determined, thanks to the linear material law, by applying a simple rule of three. Since σ_V used in calculations is taken when applying F_{proof}, an additional multiplication with S_{pf} is necessary. Thus LF_{el} is defined as

$$LF_{el} = \frac{R_e S_{pf}}{\sigma_V}. \tag{6}$$

LF_{pl}, on the other hand, is defined as the load factor acting on the investigated component when $\varepsilon_{pl} = 15\,\%$ is reached for the first time in the critical area of the weld or the eyehook. The decision to limit the maximum allowable plastic strain to $\varepsilon_{tol} = \varepsilon_{pl} = 15\,\%$ is based on a company internal criterion in order to have sufficient safety up to the materials $A_{\%}$. The tolerable equivalent stress according to the plastic form factor is defined as $\sigma_{tol} = K_p R_e$.

IV. RESULTS

Due to a confidentiality clause, no concrete values for stresses or yield strength's are given. Additionally, all illustrations are depicted with normalized legends.

A. Component

The submodel's critical area under investigation is the eyehook, respectively the upper part and the eye itself, highlighted in Figure 7. The reason for that is that the lifting equipment only directly acts on the upper part while the lower area is already influenced by the weld, which does not allow an influence-free evaluation of the eyehook itself.

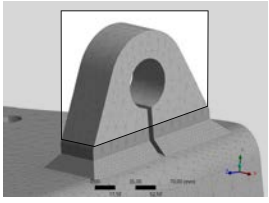
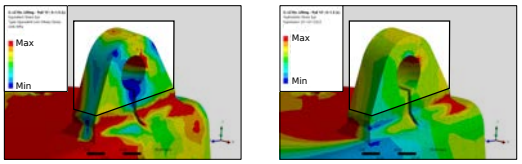


Fig. 7. Investigated and evaluated upper part of the lifting eyehook.

For the calculations necessary to determine K'_p the average equivalent stress $\sigma_{V,avg}$ and the average hydrostatic stress $\sigma_{H,avg}$ of all three concepts is used. In Figure 8 σ_V and σ_H are shown with the focus on the detail of Figure 7.



(a) (b)

Fig. 8. σ_V (a) and σ_H (b) taken from the structural stress model.

Table IV below represents all essential results necessary to calculate K'_p. The notch stress concept is evaluated separately since ε_{ref} is defined differently in [9].

TABLE IV
EVALUATION OF THE EYEHOOK ACCORDING TO [9].

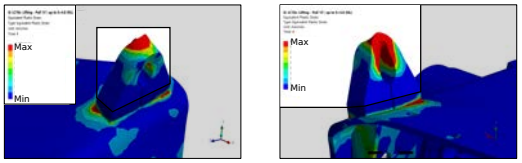
	Nominal & Structural Stress	Notch Stress
h	0.39	0.39
ε_0	5.00 %	5.00 %
ε_{ref}	22.00 %	17.00 %
ε_{tol}	20.52 %	15.40 %
K'_p	11.03	9.55

According to $K'_p \geq 9.55$ the eyehook is capable of withstanding over nine times the load which is necessary to reach $\sigma_V = R_e$. Nonetheless, K'_p is the upper limit which is rarely decisive for determination of n_{pl} [12].

For the evaluation by the plastic form factor K_p it is not necessary to differentiate between all three weld modelling concepts, as these have no actual influence on the upper part of the eyehook. Through incorporation of formula (6) into formula (5) it is possible to obtain

$$K_p = \frac{\sigma_V}{R_e} \frac{LF_{pl}}{S_{pf}}. \tag{7}$$

The high plastic strains and deformations, see Figure 9, occurring in the non-linear simulations are very local effects. A reason for this is the type of force application, which seems to be too conservative. For this reason, and due to the fact that these extreme strains and deformations will not occur in reality, $LF_{pl} = 4.00$ can be guaranteed nevertheless [13].



(a) (b)

Fig. 9. Plastic strain at the eyehook after applying loads up to $S_{bf} = 4.00$.

By inserting $LF_{pl} = 4.00$ and $S_{pf} = 1.50$ as well as the characteristic values for R_e and σ_V into formula (7) it is possible to determine K_p. In this case $K_p = 4.69$, which means the eyehook is capable of withstanding four-point-sixty-nine times the load which is necessary to reach $\sigma_V = R_e$.

B. Weld

The weld concepts are examined individually as each concept leads to different results, due to the different level of detail in the modelling process. For the evaluation, mainly the weld toes between the upper and lateral cover plates and the weld itself are considered. Due to the load case dealt with, the area highlighted in Figure 10 is of great interest, as the greatest stresses occur here.

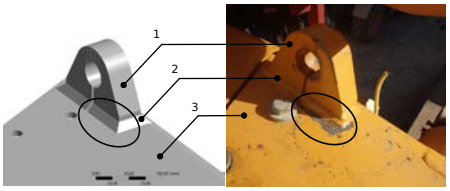


Fig. 10. Critical area of the fillet weld at the front lifting point with eyehook (1), fillet weld (2) and upper cover plate (3) shown in Ansys (left) and on the real machine (right) [13].

The evaluation of the weld concepts according to the elongation criterion leads to the results in Table V below. The nominal stress concept has the highest value of all concepts with $K'_p = 10.00$. Beside that it has also the lowest σ_V when F_{proof} is applied. The reason for this is that there is no weld which can influence the stress distribution. The structural stress concept is the most conservative of all concepts with $K'_p = 5.44$. Since it is a fillet weld, an additional weld factor of $\alpha_W = 0.8$ must be considered to determine $\sigma_{V,tol}$. Thus $\sigma_{V,tol}$ is defined as $\sigma_{V,tol} = K'_p R_e \alpha_W$. The result of the notch stress concept is, despite the highest stresses, with $K'_p = 8.82$ between the other two concepts. According to the evaluation by applying the elongation criterion, the weld with modelling based on the structural and notch stress concept would already fail when applying the F_{proof} and only the nominal stress concepts would prevail. This can be stated, since $\sigma_{V,tol}$ determined with K'_p is lower than σ_V reached by the linear FEA.

TABLE V
EVALUATION OF THE WELD CONCEPTS ACCORDING TO [9] AND LINEAR FEA

	Nominal Stress	Structural Stress	Notch Stress
h	0.54	0.59	0.47
ε_0	5.00 %	-	5.00 %
ε_{ref}	22.00 %	-	17.00 %
ε_{tol}	16.89 %	5.00 %	13.15 %
K'_p	10.00	5.44	8.82

The plastic form factor K_p for each weld modelling concept is determined by substituting the characteristic values, obtained by the linear and non-linear FEA, into formula (7). The more detailed the modulation becomes, the smaller LF_{pl} gets, which means the earlier $\varepsilon_{pl} = 15\,\%$ is reached. As with the elongation criterion, mainly the weld toes in the critical area are considered. The nominal stress concept reaches the highest LF_{pl}. It can even be assumed that $LF_{pl} \geq 4.00$, since $\varepsilon_{pl} \leq 15\,\%$. This leads to $K_p = 9.00$, which means that

stresses of up to nine times of the material's R_e are allowed before failure at the weld toe must be assumed. In Figure 11 the plastic strain at the area where the weld would normally be is depicted.

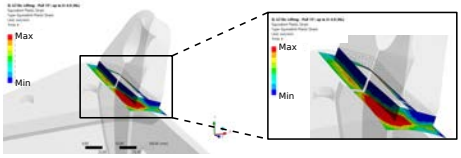
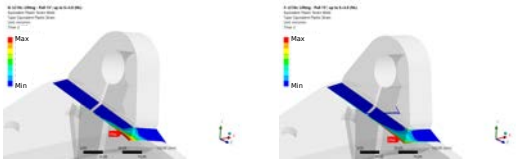


Fig. 11. Plastic strain ε_{pl} of the nominal stress concept at the weld area.

The structural as well as the notch stress concept turn out to achieve relatively small plastic load factors of $LF_{pl} = 1.70$ and $LF_{pl} = 1.20$. These indicate that a failure of the weld is already possible at a load in the proximity of F_{proof}. This does not to correspond to reality, as it has also been proven by [13]. The resulting form factors of $K_p = 6.90$ and $K_p = 9.23$ suggest a high maximum load limit, but this is not the case. The reason for this is that the elastic load limit L_{el} is relatively low, i.e. the condition $\sigma_V = R_e$ is fulfilled relatively early. This is confirmed by the high equivalent stresses σ_V of the two concepts although only F_{proof} is applied. These stresses should not be used as the sole basis for evaluating the weld.



(a) (b)

Fig. 12. Plastic strain ε_{pl} at the weld based on the structural (a) and notch stress concept (b) after an applied load factor of $LF = 2.00$.

In Table VI an overview of all aforementioned K_p's is given.

TABLE VI
WELD EVALUATION BASED ON THE NOMINAL, STRUCTURAL AND NOTCH STRESS CONCEPT.

	Nominal Stress	Structural Stress	Notch Stress
S_{pf}	1.50	1.50	1.50
LF_{pl}	4.00	1.70	1.20
K_p	9.00	6.90	9.21

C. Result Overview & Explanation

In Table VII an overview of all determined values for K_p' and K_p is given. Besides that the plastic support number n_{pl}, which is based on formula (1), is also stated. As mentioned in [10] and [12], in almost all cases the plastic form factor K_p is the decisive criterion for determining n_{pl}. For the eyehook, $n_{pl} = 4.69$ is the same for all three concepts, since the modelling of the weld has no influence on the eyehooks load capacity. The slightly different values of K_p' are due to the different specifications given in [9]. More interesting, however, is the evaluation of the weld and its different modelling concepts. The application of the elongation criterion shows different values for K_p' for all three concepts. The plastic form factor K_p is similar for the nominal and notch stress concept but they may not be compared directly. The same is true for the structural stress concept, which stands out because K_p' is significantly smaller than K_p and decisive for the definition of n_{pl}. However, as mentioned before, the structural and notch stress concept must be viewed very critically in this case.

TABLE VII
DETERMINED VALUES FOR K_p', K_p AND THE DERIVED n_{pl}

	Nominal Stress	Structural Stress	Notch Stress
Eyehook			
K_p'	11.05	11.05	9.55
K_p	4.69	4.69	4.69
n_{pl}	4.69	4.69*	4.69
Weld			
K_p'	10.00	5.44	8.82
K_p	9.00	6.90	9.21
n_{pl}	9.00	5.44*	8.82

*additional weld factor α_W not included

In order to make a statement about the compliance with the applicable standards one can not directly use the determined n_{pl}. The reason for this is that n_{pl} only provides a statement about how far the investigated part or component can be loaded beyond the material's R_e.

V. General Evaluation

This chapter is dedicated to the assessment and evaluation of the FE simulations. It aims to provide a clear and structured overview of the most important results. The focus is not only on the explicitly treated component from section II-C and IV, but takes into account all investigated components and structures.

A. Plastic Strength Reserves

As the investigations have shown, the plastic strength reserves of industrial steel are considerable. Regardless of whether S_{pf} or S_{bf} specified in the standards are fulfilled, the elastic load limit L_{el} can be exceeded many times over by L_{pl} without having to assume component failure. If the elongation criterion, which is rarely decisive for determining the plastic support number n_{pl}, is ignored, the non-linear FEA consistently delivers high results for n_{pl} which indicate massive plastic material reserves. Nonetheless, the accuracy of the results should always be questioned and considered more critically. The reason for this is that the mathematical representation of a real problem can only be made with simplifications and assumptions, which can falsify the results. This can already have considerable effects in linear FE-simulations, but in non-linear FE-simulations such wrong assumptions and simplifications are even worse and can hugely influence the results.

B. Compliance with Standards

The applicable standards specify a proof factor S_{pf} and breaking factor S_{bf} that the component or structure has to fulfil. Since the focus of this work is set on plastic material behaviour, the breaking factor S_{bf}, which allows plastic material behaviour, plays an essential role. The plastic support number specified by the standard $n_{pl,ISO} = S_{bf}/S_{pf}$ is determined by the ratio of both factors. In Table VIII all $n_{pl,ISO}$ to be fulfilled are listed.

TABLE VIII
PLASTIC SUPPORT NUMBERS $n_{pl,ISO}$ SPECIFIED BY THE STANDARDS.

Operation	$n_{pl,ISO}$
Retrieval acc. to [3] and [7]	1.50
Lashing acc. to [2]	1.60
Lifting acc. to [2]	2.67

With the help of Table VIII, it can be checked whether the examined components and structures meet the standards or not. For this purpose, the LF_{pl} obtained from the non-linear FE-simulation is used, since it corresponds to the factor that the component can be loaded beyond the materials R_e. Since a tolerable plastic strain of $\varepsilon_{tol} \leq A_{\%}$ is already considered during the evaluation process, there is no need to assume failure. If $LF_{pl} \geq n_{pl,ISO}$ the standard is met, if $LF_{pl} < n_{pl,ISO}$ the standard is not met. In Table IX this compliance with the applicable standard is shown for the investigated front lifting point of the crawler tractor PR746.

TABLE IX
COMPLIANCE OF THE CRAWLER TRACTOR'S FRONT LIFTING POINT WITH ISO-15818.

	Nominal Stress	Structural Stress	Notch Stress
Eyehook			
LF_{pl}	4.00	4.00	4.00
Compliance?	✓	✓	✓
Weld			
LF_{pl}	4.00	1.70	1.20
Compliance?	✓	x	x

As already mentioned before, the structural stress and notch stress concept would indicate that the lifting point does no fulfil the applicable standards criterion. But as it has also been proven by a real structural test, this is not true. Once more this shows that the results of non-linear FE-simulations and the applied weld modelling concepts have to be viewed critically. Table X below lists all investigated components, applicable standards and their compliance.

TABLE X
COMPLIANCE OF ALL INVESTIGATED COMPONENTS WITH THEIR APPLICABLE STANDARDS.

Component	Machine	Standard/s	Compliance
Counterweight	PR726	[2], [3]	✓
Counterweight	PR746	[2], [3]	✓
Counterweight	PR766	[2], [3]	✓
Fall Protection	PR776	[14], [15]	✓
Lifting Point	PR746	[2]	✓
Lifting Point	PR736	[2]	✓
Lifting Point	T55-7	[2]	✓
Lashing Point	T55-7	[2]	✓

C. Applicable Methodology

In order to avoid time-consuming calculations and simulations for similar components in the future, n_{pl} determined from these investigations should be able to be converted and applied on similar components. This methodology is briefly explained below. With the linear FEA the elastic load limit L_{el} is determined by

$$L_{el} = \frac{\sigma_V}{R_e} F_{abs} \tag{8}$$

where F_{abs} is the applied load. The next step is to determine the plastic load limit L_{pl} with

$$L_{pl} = L_{el} K_p = L_{el} n_{pl}. \tag{9}$$

The final step is then to compare L_{pl} with the breaking force F_{break} given by the standard applicable. If the condition $L_{pl} \geq F_{break}$ is fulfilled, it can be assumed that the standard is met. Additionally, one should calculate $\sigma_{V,tol}$ by inserting n_{pl} and the material's R_e into formula $\sigma_{V,tol} = n_{pl} R_e$. The resulting $\sigma_{V,tol}$ gives an indication about the tolerable stresses that can be reached in the linear FEA.

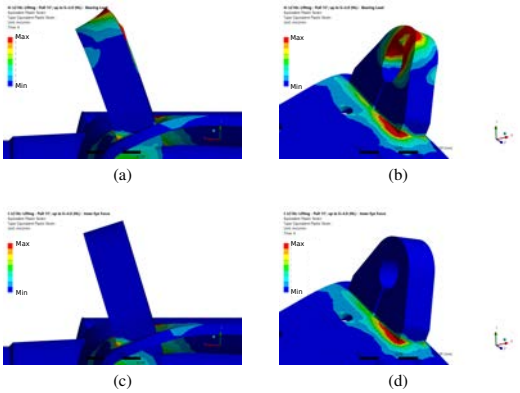
D. Modelling of Welds

It is hardly possible to make a generally valid statement about when which concept should be applied. Nonetheless, some statements about the modelling of welds in general can be made from the obtained results. From a purely economic, time-consuming and cost-technical point of view, the nominal stress concept makes the most sense. It leads to the fastest computation times, has the lowest modelling effort and consistently delivers results, which allow a good assessment of the component's actual behaviour. The structural stress and notch stress concept, on the other hand, entail a much higher modelling and computational effort. Furthermore, the results must be considered more critically, since singularities can occur due to both the meshing and the geometry itself, which can lead to a falsification of the results. Nevertheless, these concepts can help to identify critical areas of the weld and thus show where possible cracks and fractures can occur. The usefulness of accurate weld modelling is highly dependent on the component and load case. For small welded components such as the cab lifting point or the personal fall protection, the application of the structural stress or notch stress concept is useful to obtain more accurate results. For large components and structures such as the counterweights or the front lifting point examined above, the effort of modelling the weld more accurately is not always worthwhile.

E. Improvements of Simulation Inaccuracies

In-depth investigations often reveal points that need to be improved. In the course of this work it became apparent that certain modelling approaches that are used for linear FE-simulations should not be transferred to non-linear FE-simulations. The force application at the eyehook can be taken as a concrete example. The application of force used leads to a result falsification when non-linear material behaviour is considered. The deformations reached do not occur in reality,

which is also confirmed by [13]. Due to this, two alternative methods of force application were modelled and simulated, see Figure 13 below. These are clearly less destructive than the initial method, shown in Figure 9, and thus can be brought closer to the results of the real structural test, see Figure 14.



(a) (b)

(c) (d)

Fig. 13. Result comparison of the two alternative methods for force application at the eyehook, with little to no effect on the weld area.

(a) (b)

Fig. 14. Results of the real structural test show that the actual deformations differ from the results provided by the non-linear FE-simulations.

VI. Conclusion

The original objective of this work, an investigation of the non-linear material behaviour under loads surpassing the materials yield strength and the utilisation of this behaviour to meet the applicable standards, was successfully investigated and confirmed. As it turned out, some of the components examined have immense plastic reserves, which go far beyond the standards. The additional investigation of different concepts for modulation of welds showed that the these results should be viewed critically and with caution. This is because, depending on the component being investigated, more accurate modelling does not always provide more accurate results. Two questions remain open, namely whether and to what extent these simulations correspond quantitatively with reality and to what extent the results obtained can be converted to similar components and machine types. Nevertheless, these numerical investigations have resulted in some new findings and understandings, which in turn have opened up new possibilities and perspectives. These can in turn be taken into account and consulted for future investigations and help to guarantee maximum safety for user and machine, even in extreme situations.

Acknowledgement

The main author would like to thank Dr. Claus Lamprecht, technical calculation manager and supervisor at Liebherr-Werk Telfs, and Dr. Franz-Josef Falkner, MCI supervisor and lecturer at the MCI, for the support during the writing of this master thesis. A big thank you also goes to the author's fellow students with whom he had the privilege of experiencing a excellent and unique collaboration throughout the entire study period. And finally, special thanks to the author's family, who made this valuable life stage possible.

References

[1] International Organization for Standardization, "ISO 6165:2012(E): Earth-moving machinery – Basic types – Identifications and terms and definitions," Geneva, CH, 2012.

[2] ——, "ISO 15818:2017(E): Earth-moving machinery – Lifting and tying-down attachment points – Performance requirements," Geneva, CH, 2017.

[3] ——, "ISO 10532:1995(E): Earth-moving machinery – Machine mounted retrieval device – Performance requirements," Geneva, CH, 1995.

[4] LWT – Technical Documentation Department, "Operator's manual: Crawler dozer," Telfs, AUT, 2019.

[5] A. Hobbacher, "Fatigue design of welded joints and components," Cambridge, ENG, 1996.

[6] K. Nagarajan, "Design of Structures: Lesson 7 – Welded Connection," Coimbatore, IND, 2021.

[7] European Committee for Standardization, "EN 1459:2017(E): Rough-terrain trucks - Safety requirements and verification - Part 1: Variable-reach trucks," Brussels, BEL, 2017.

[8] W. Rust, "Nichtlineare Finite-Elemente-Berechnungen mit ANSYS Workbench: Strukturmechanik: Kontakt, Material, große Verformungen," Wiesbaden, GER, 2020.

[9] R. Rennert, E. Kullig, M. Vornwald, A. Esderts, and D. Siegele, "FKM-Richtlinie: Rechnerischer Festigkeitsnachweis für Maschinenbauteile aus Stahl, Eisenguss- und Aluminiumwerkstoffen," Frankfur/Main, GER, 2012, 6., revised edition.

[10] M. Fiedler, I. Varfolomeev, and M. Wächter, "Richtlinie Nichtlinear: Rechnerischer Bauteilfestigkeitsnachweis unter expliziter Erfassung nichtlinearer Werkstoffverformungsverhaltens," Frankfur/Main, GER, 2016.

[11] H. Neuber, "Kerbspannungslehre: Grundlagen für genaue Spannungsberechnung," Berlin/Heidelberg, GER, 1937.

[12] Einbock, Stefan, "Die Neuber-Regel. Oder: warum plastische Verformungen gut sind und wie Sie Bauteile mit der plastischen Stützzahl höher auslasten," 2020, Last seen on 24.06.2021. [Online]. Available: https://www.einbock-akademie.de/die-neuber-regel-und-die-plastische-stuetzzahl-einfach-erklaert

[13] LWT – Technical Office & Technical Test Department, "Versuchsanweisung-Nr.02202851," Telfs, AUT, 2017.

[14] European Committee for Standardization, "EN 795:2012(E): Personal fall protection equipment – Anchor devices," Brussels, BEL, 2012.

[15] ——, "EN 363:2019(E): Personal fall protection equipment – Personal fall protection systems," Brussels, BEL, 2019.

Nico Schuchter is a master student at MCI Innsbruck/Austria. He received his bachelor's degree in mechatronics, major mechanical engineering in 2019 from the MCI, where he is currently working toward his master's degree in mechatronics & smart technologies. He completed his bachelor and master thesis's in collaboration with Liebherr-Werk Telfs (LWT), with focus on design and numerical simulation of different machine components for earth-moving machinery.

Experimental Determination of the Bending and Torsional Stiffness of an Alpine Ski.

Hannes Alexander Schwendinger and Franz-Josef Falkner (supervisor)

***Abstract*—The development of alpine skis is currently strongly dependant on the designers and ski-testers empirical knowledge. Though, geometrical and mechanical properties are crucial for skiing performance. A ski is layered from various materials, such as wood, metals, composites and other polymers to a sandwich system. This layout provides a complex mechanical structure, and therefore, the distribution of bending and torsional stiffness is not readily available. Anyhow, it is still possible to describe bending and torsion with the basic mechanical concepts. To determine the stiffness values of a ski, knowledge of deformations and the acting loads allow an indirect measurement of those parameters. In order to get rid of the influence of measurement uncertainties, the mechanical laws are not applied directly. A virtual bending line / torsion line is fitted on the measurement data, where the stiffness values are the parameters to be optimized. The outcome of this procedure provides qualitative and quantitative distribution of the bending and torsional stiffness. The result of the bending stiffness is very promising, whereas the outcome for the torsional stiffness has a chance for improvement, especially regarding the mechanical measurement.**

***Index Terms*—Alpine Ski, Measurement of Bending and Torsional Stiffness, FEM Validation, Optimization Procedure, Composite Materials.**

I. Introduction

ALPINE skiing is one of the most popular winter sports in the Austria [1]. Therefore, also the need of the related gear is high. In 2018/19, around $430\,000$ pairs of ski where sold in the Austrian shops. That is why ski manufacturers, such as *HEAD*, are steadily developing new alpine skis. Despite the big market, this procedure is still quite conventional. Skis are mostly designed based on empirical knowledge of the engineers and ski-testers, although today's possibility of simulation and design are highly advanced. The performance of alpine skis is not as empirical as the current development procedure. It is strongly linked to the geometry and the mechanical properties, such as bending stiffness (BS) and torsional stiffness (TS) [2], [3]. While the geometry is available for the designing process, the stiffness values are not. The current procedure for developing new skis happens iteratively. Engineers are designing and manufacturing a prototype, which is then accessed by a test skier. Since it is most of the time not possible to immediately design a perfect ski, there are several loops to improve the prototype. With the knowledge of the stiffness properties of developed skis, more descriptive methods can be applied for the development. Which means that the mechanical parameters can be specifically used. In this way it is most likely possible to speed up the designing process, save costs and even further improve alpine skis. Also the validation of finite elements method (FEM) simulations would be possible.

A. State of the Art

HEAD does several measurements to check the quality of their products, such as fatigue strength, breaking and ageing tests. The products provide highest quality. The mechanical properties are also investigated. The standard ISO 5902 regulates a "flex"-measurement regarding bending and torsion of alpine skis. Those measurements do not have to fulfill any boundaries or given values, they just have to be carried out. Resulting BS and TS values are just averaged and not really precise. The results are not used for the development procedure.

Truong et al. developed a method to measure the distribution of BS and TS of alpine skis. They make use of the mechanical relation between deformation, a load and the stiffness. Therefore, the ski is deformed in a combined way with a force and torque load. The curvature of the ski is measured in both horizontal directions. With the knowledge of load and deformation the BS and TS can be determined [4].

The Canadian company *Sooth Ski* developed a device for the measurement of BS and TS distribution as well as the geometry of alpine skis. The measurement is as well based on deformation. The used sensors are precise lasers. More details about the data analysis and stiffness computation are not mentioned. The values for the BS and TS are averaged for $100\,\mathrm{mm}$ segments [5].

Wikerman developed methods to measure the BS and TS as well. The BS is determined by measurement of bending-line (BL) of a deformed ski with image processing and measurement probes. A constant bending moment is applied to the ski. The ski length is discretized to finite elements (FE) and the BS is computed by a least squares minimization. To measure the TS, the ski is loaded with a constant torque. To record the torsion angle, lasers are attached to the ski, pointing on a board to visualize the rotation angle. Again, the length of the ski is discretized to FE to compute the stiffness [6].

B. Motivation

Currently carried out stiffness measurements at *HEAD* are not usable for development of alpine skis. Anyhow, the knowledge of BS and TS distribution can bring the development procedure of alpine skis to a higher level. Speeding up the designing procedure, saving costs and further improvement of alpine skis are just some benefits. Another important aspect is the validation of FE models, which are currently still in a development phase. Also the comparison of already existing ski-models would be possible on a more formal way in comparison to an empirical opinion of ski testers.

C. Objective

A qualitative and quantitative measurement, respectively determination, of the BS and TS distribution for alpine skis. The maximum measurable ski length should be around $2155\,\mathrm{mm}$, which is equal to the longest downhill ski. Maximum width of the ski should be around $160\,\mathrm{mm}$, which can be the case for freeride skis [7]. The measurement needs to happen without destroying the skis and should happen in an adequate time span, without the need of adjustments at the ski itself. It should be simply put into the device to be measured. A deformation of the ski for the bending test should be initialized by a force of maximum $500\,\mathrm{N}$ and result in a maximum deflection of $100\,\mathrm{mm}$. The maximum torque acting on the ski should be around $75\,\mathrm{N\,m}$, leading to a maximum torsion angle of $30°$. Determined values for the BS and TS should have a precision of $\pm 5\,\%$ of their maximum value.

II. Theoretical Background

In order to meet the desired skiing performance, alpine skis need to have a unique structure. The needed stiffness values of skis are reached with the layering procedure. Neglecting the complex structure of the ski, the basic mechanical relations are still valid. For the determination procedure, some numerical procedures are needed.

A. Structure of an Alpine Ski

Modern skis are available in countless shapes and styles to fit users demands. Beside the different geometries also the BS and TS are strongly variable and important for the performance [2], [3]. *HEAD* provided two test skis for the development procedure of the experiment. Those are standard all-round skis with a length of $1600\,\mathrm{mm}$ and a centre width of $70\,\mathrm{mm}$, shown in Fig. 1.

Fig. 1. Exemplary alpine ski, in this case a test ski for BS and TS measurement.

Modern alpine skis are layered to a sandwich system. The core is usually a conical wood board covered with various materials, such as glass reinforced plastic (GRP), carbon reinforced plastic (CRP), metals rubber and other polymers. This combination of highly different materials, regarding the mechanical properties, makes the ski a complex structure. Isotropic, anisotropic and orthotropic materials are compounded to generate the needed properties. An exemplary structure is shown in Fig. 2.

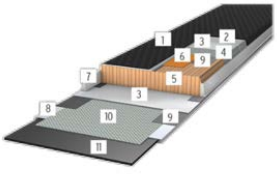


Fig. 2. A *HEAD* WC CAP Construction: 1 Top Sheet, 2 Glass Fleece, 3 Titanal Layer, 4 Glass Fibre, 5 Wood Core, 6 Graphene, 7 Side Wall, 8 Edge, 9 Dampening Layer, 10 Reinforced Fibre Glass and 11 Base. [7]

B. Bending Beam

Despite the complex structure of alpine skis, the mechanical principle to describe the bending is rather simple. The conventional Euler-Bernoulli beam theory is applied, because the shape of a ski can be described by a beam. It is assumed that rather small bending deflections result in a plane cross-section which also stays perpendicular to to the beam axis. For a small deflection, the rotation angle of the cross-section can be approximated with $\frac{\mathrm{d}w}{\mathrm{d}x}$ which finally leads to the ODE for a slender beam [8]

$$M(x) = -EI(x)\frac{\mathrm{d}^2 w}{\mathrm{d}x^2} = -EI(x)w''. \tag{1}$$

Since the BS is the unknown parameter and is additionally not constant the formulation for a ski reads as

$$EI(x) = -\frac{M(x)}{w''}. \tag{2}$$

For this inhomogeneous beam the BS, $EI(x)$ has to be handled as one parameter. The ski is basically simplified to a planar single-span beam, schematically shown in Fig. 3.

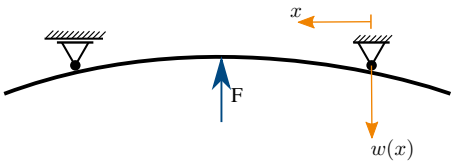


Fig. 3. The Ski is handled as a usual planar single-span beam. It is loaded with a single force, introducing a bending moment. Deflection of the beam is denoted as $w(x)$.

To prove the concept of the mentioned theory, a composite beam example with high similarity to a ski is tested at a three-point bending test. The difference is a non-changing cross-section over the beams length. The BS is calculated by

$$EI = \sum_{i=1}^{n} \frac{E_i b_i h_i^3}{12} + b_i h_i (z_i - z_n)^2 \tag{3}$$

to compare the result of the bending test with the theory. In comparison to a homogeneous cross-section, the composite one has a shifted neutral axis

$$z_n = \frac{\sum_{i=1}^{n} A_i E_i z_i}{\sum_{i=1}^{n} A_i E_i} \tag{4}$$

calculated with the parallel axis theorem [9]. In this case E is the Young's modulus, b the width, h the hight and z the centre point of each layer i.

C. Torsional Beam

In analogy to the bending beam, the torsional beam is also described by the standard differential formulation, reading as

$$\frac{d\theta}{dx} = \theta' = \frac{M_t}{GI_t(x)}. \tag{5}$$

The ski in this case is considered at cantilever beam with a torque, applied to the free end, shown in Fig. 4.

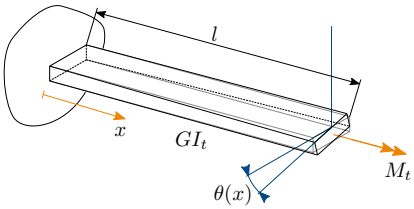


Fig. 4. Torsion of a beam suspended as a cantilever with a length l and a TS GJ_t. The acting torque M_t rotates the beam with an angle $\theta(x)$.

A theoretical calculation of TS for a composite beam is not possible analytically. Torsion of a composite beam will always lead to a combined load and tend to have a warping cross-section [10].

D. Optimization Algorithm

Due to problems with a direct application of the mechanical laws to the measured data, a least squares minimization procedure is used to search for the BS and TS. Within MATLAB, the *lsqnonlin(...)*-function is used to solve the problem. This function is based on a trust region method (TRM), which is basically an iterative procedure to search for the minimum of a function [11]. Around a starting point, a *trust region* is formed. Here, the function is approximated with a second order Taylor series approximation. With the minimum of the region, a new point is defined as well as a new trust region around it. Again the function is approximated and minimized for the new region. This procedure is repeated as long as a stop criterion is reached, such as a minimum solution change [12].

To solve a boundary value problem (BVP) within the optimization procedure, the MATLAB-function *bvp5c(...)* is used. Here, finite differences (FD) is exploited to solve differential equations while approximating derivatives [13]. FD is a rather old methodology, still is is a quite simple but powerful tool to solve differential equations [14], [15]. A grid function is used to discretize the domain, which means that the function is only defined at discrete points. All needed derivatives are also available at those points, calculated by divided differences and Taylor expansion [15].

III. Method

A direct measurement of BS and TS is not possible. With the use of the mentioned mechanical relations an indirect measurement is carried out. Meaning that the stiffness values can be determined by known load application and the resulting deformation. In order to load and measure alpine skis, a device, the so called ski test bench (STB), is designed and manufactured. The recorded data needs to be conditioned and evaluated to compute the desired stiffness distributions.

A. Ski Test Bench

The STB, shown in Fig. 5, basically consists of an aluminium frame to which all needed parts and modules are mounted. A control cabinet is attached within the frame, housing the data acquisition (DAQ) board, a controlled power source as well as the control unit for a linear actuator. Supports for the ski suspension are mounted to the left and right end. A bending module is attached to apply a bending moment as well as a torsion module can be used to apply torque. The measurement head is attached at the top of the frame to record the introduced deformations. All parts, beside the control cabinet, are movable. Linear scales are attached to simplify positioning.

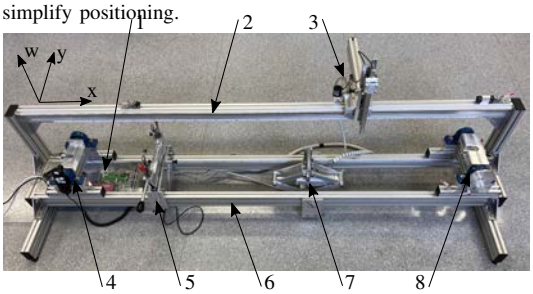


Fig. 5. Ski test bench setup: 1 control cabinet, 2 linear guide, 3 measurement head, 4 front suspension, 5 torsion module, 6 Item frame, 7 bending module and 8 rear suspension.

Special supports are designed and manufactured to ensure the suspension of the skis. They consist of pedestal bearings, ensuring a free rotation of the ski at the support points. The front suspension, acting as the floating bearing does not allow a clamping of the ski, where the rear suspension has an integrated toggle clamp to block horizontal movement. The rear support also provides the possibility to block its rotation. In this way it can also be used for the torsional test, which requires a cantilever beam configuration. Both suspensions are designed in a way, that the ski is directly positioned at the rotation point of the bearing.

To apply a force to the ski, resulting in a bending deformation, the bending module is used. It basically consists of an electrical linear actuator placed in a scissor mechanism. A force sensor with a range of 0 to $500\,\mathrm{N}$ is mounted to the top of the module measuring the load on the ski. The mechanism provides a lifting hight of $120\,\mathrm{mm}$ and a maximum force slightly extending the sensors range. It is slide-able along the frame to adjust the desired force introduction position.

For the torque application, a torsion module is designed and manufactured. The main idea is to fix the ski in the module in a way, that the centre-point of the ski-cross-section is directly at the rotation-axis. The mechanism is schematically shown in Fig. 6. A rotating plate connects the ski to three ball bearings, which allow only one degrees of freedom (DOF), namely the rotation θ. The torque is applied with a lever. Once the ski is twisted, the mechanism is blocked by two spring plungers, which translate the torque to a force sensor.

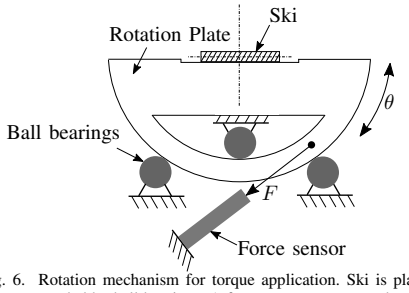


Fig. 6. Rotation mechanism for torque application. Ski is placed on rotary plate suspended by ball bearings. A force sensor measures the applied torque.

To record the BL, respectively torsion-line (TL), a measurement head is attached to the top of the device. It is connected to the frame with a linear guide, allowing a horizontal movement in the x-direction. End stops allow adjustment to different ski lengths. The vertical displacement, w-direction, is measured with a highly precise optical linear encoder. It provides a maximum resolution of $2\,\mathrm{nm}$ [16]. A probe is fixed to the sensor, guided by a linear conduction to sense the ski surface. The mechanism can be arrested in its initial position by a spring plunger. For the recording of the horizontal position, x-axis, a rotary encoder is attached to the measurement head. With a gear-wheel it is connected to a gear-shaft placed at the frame. With an X4 counting scheme, it is possible to generate a horizontal resolution of $0.02\,\mathrm{mm}$. To record the TL, it is necessary to measure two traces on the ski-surface for an angle calculation. For this, the measurement probe can be shifted in y-direction. Two end-stops ensure precise positioning for the measurements.

B. Bending Stiffness

The ski is suspended with the front and rear suspension as single-span beam. Running surface on the top. Shown in Fig. 7. The bending module is placed in the centre of the ski. Since the ski commonly provides pretension, it is not flat. The theory behind a bending beam is based on a straight beam. To avoid this problem a referenced measurement approach is carried out. Therefore, the ski is slightly bended with the bending module. Which means that the applied displacement is around 10 to $20\,\mathrm{mm}$. Now, the measurement head, moreover the probe, is placed on the running-surface of the ski at the rear suspension. The probe is moved over the length of the ski to the front suspension to record a reference BL. Simultaneously, the applied force is measured. Now, the ski is further deformed, shown in Fig, 7. The bending module is, depending on the ski, lifted between 30 and $60\,\mathrm{mm}$. Again the BL and force is recorded. With the calculation of the differences between both measurements, the final BL is calculated.

Fig. 7. Bending test: A ski in the deformed measurement configuration within the STB.

The DAQ happens with MATLAB. A graphical user interface (GUI), shown in Fig. 8, is programmed for the measurement procedure. By simply selecting a measurement duration, a filename and pressing a button, the measurement can be started. After the given time is up, the recorded BL is plotted. Applied force, torque and displacement is visualized. Recorded data is saved to the selected filename.

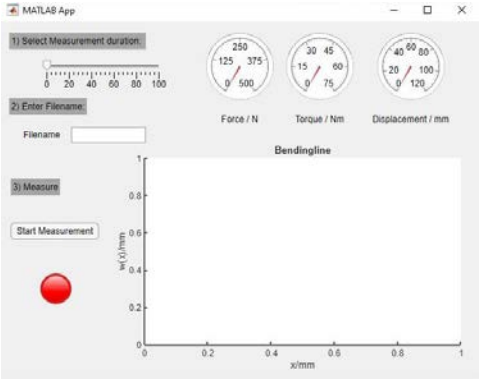


Fig. 8. MATLAB measurement GUI.

The recorded data needs to be conditioned for the determination of BS. First of all, the offset, coming from the reference position of the vertical encoder needs to be reduced. Due to the different resolution of the horizontal and vertical sensors, one horizontal sample provides several vertical samples. This data is reduced by averaging. Although the signals are both digital, a very small ripple occurs, shown in Fig. 9. This may comes from the measurement itself or from the mechanism. For the computation of the BS, this leads to problems. The BL needs to be filtered.

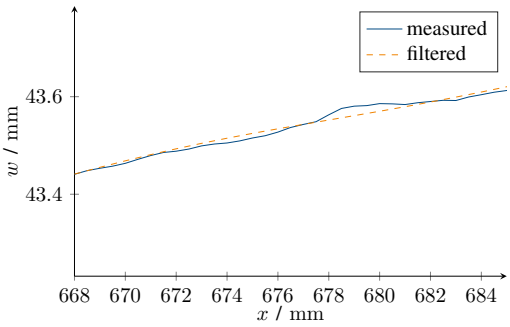


Fig. 9. Close-up of BL data before and after filtering.

The approach in this case can not be a conventional low-pass-filtering. The given data is not time dependant. A filtering procedure over the wavelength is applied. As the ripple provides a wave length, which is mechanically not possible, it needs to be excluded. To do so, a wave spectrum analysis is carried out. For this, the fast Fourier transform (FFT) is applied. By investigations of the occurring wave lengths, respectively the corresponding wave-numbers in Fig. 10, it can be seen that they widely extend the plausible range. To reduce this unwanted ripple, a maximum wavelength is defined with $20\,\mathrm{mm}$. Everything lower is cut off. The BL is than recreated by use of the inverse fast Fourier transform (IFFT). The finished BL is shown in Figure 11 and the comparison of the filtered and unfiltered signal is shown in Fig. 9.

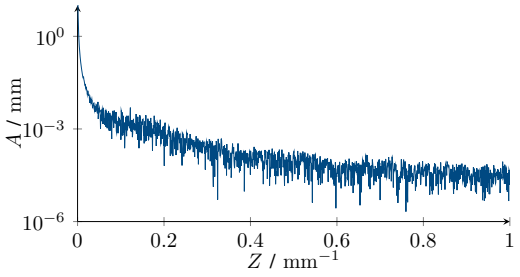


Fig. 10. Wave-spectrum of the BL.

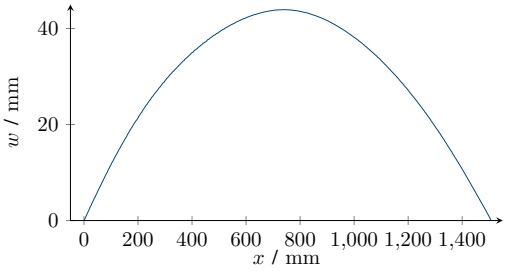


Fig. 11. BL generated out of subtraction of reference measurement from deformed measurement.

The bending moment is defined as piecewise polynomial. The conditioned BL, the force data and the bending moment data is saved to the current folder. For the determination of the BS, a separate MATLAB script loads the conditioned data. Since the BL still has some uncertainty in it, it is not possible to directly apply the mechanical law. Using the second derivative of $w(x)$ would lead to a high unstable result, because the already very small fluctuations on the BL will be amplified. That is why the approach is to fit a virtual BL to the measured data, as shown in Fig. 12. The main idea is to set up a least squares minimization procedure, reducing the error between virtual bending line and measured data. The optimization parameter is the BS. Within MATLAB, the function *lsqnonlin(...)* is used to solve this non-linear least squares minimization. As objective function, the mentioned mechanical relation is used. Since this objective function is a second order ordinary differential equation (ODE), it is solved with *bvp5c(...)*. The second order ODE must be rewritten as system of first order ODE to be solvable.

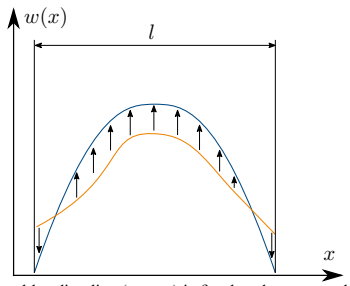


Fig. 12. A virtual bending line (orange) is fitted to the measured data (blue). The parameter which is optimized is the BS. l is the length of the ski, respectively the length of the BS determination.

For the determination of BS, the first and last $80\,\mathrm{mm}$ of the BL are neglected because the supports may influence the stiffness. For the optimization procedure, lower and upper boundary values are needed, as well as initial values. The boundaries are defined by a range, which easily extends the possible stiffness range of the ski, but limits the procedure to a certain area. The initial values are firstly inserted as a common average stiffness of a ski. At the beginning, the problem is solved for a constant distribution of BS to generate an appropriate initial value. Afterwards, the problem is solved for the distribution of BS. The difference in the solver settings is just the number of supporting points (SP) and the initial value. For the constant BS calculation, the number of SP is 1. For the distribution of BS, the rule of thumb is to use one SP for every $100\,\mathrm{mm}$ of measurement length. The regions in between the points is interpolated with a Hermitian interpolation polynomial of third order to enable a smooth transition. Once the determination of BS is carried out, the BL is recalculated from the determined stiffness and the error between the measurement and the determined BL can be investigated.

C. Torsional Stiffness

For the measurement of the TS, the ski is suspended as cantilever beam with the rear support. The torsion module is used to create a rotating support at the front end of the ski. The ski is rotated, depending on the length and stiffness, for about 10 to 30°. In order to measure the rotation angle over the length of the ski, two measurements are required. As shown in Fig. 13, the measurement probe is used to measure two traces with a defined distance d in between. For the data recording, the same MATLAB GUI is used as for the bending test (Fig. 8).

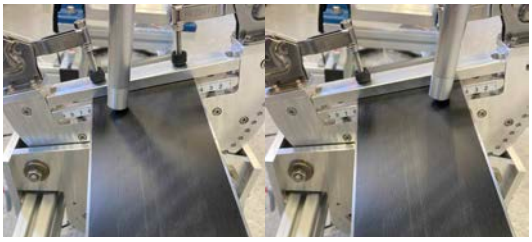


Fig. 13. Torsion test: Cross-section of the ski is rotated around its centre line. Two traces are recorded during the measurement to recalculate the torsion angle θ.

The recorded data is again conditioned before the TS computation. The initial offset to the reference position is subtracted. Whereas the different traces already provide a rotation angle at the beginning of the measurement, the initial position needs to be subtracted with respect to one of the traces. A data reduction is carried out due to the different resolution of the horizontal and vertical sensor.

Since the measurement procedure is identical to the one applied for the BL measurement, it is necessary to filter the signal again. Basically, an equivalent procedure is applied. The filtering happens over the wave-length within a wave-spectrum analysis. It has to be mentioned, that the TL does not meet one important aspect to carry out the FFT. A signal for this

analysis needs to be periodic. This also is not true for the BL, but a concatenation of several similar signals would be periodic and would have a stable and smooth transition. This is not true for the TL. A concatenation of those curves would not lead to a smooth transition, it is contrary. A sudden jump over the whole amplitude would appear. That is why the signal needs to be modified. The TL is attached to itself in a mirrored format. Which means that the curve is extended to ensure a stable transition. With the created curve, the same filtering procedure as for the BL is applied. In this case the maximum wave-length is set to $100\,\mathrm{mm}$, which was established during the application. After the filtering, the data is again reduced to the initial length. The torsion angle is calculated of the two filtered traces by

$$\theta(x) = atan\left(\frac{w_1(x) - w_2(x)}{d}\right), \tag{6}$$

where $w_1(x)$ is the vertical displacement of the first trace and w_2 of the second. The TL is shown in Fig. 14. d represents the distance between both traces. The torque is constant over the length of the ski. Nevertheless, it is defined as a polynomial to stay with the same procedure compared to the BS computation. The conditioned data is saved to the current folder.

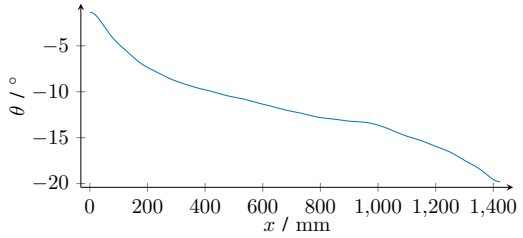


Fig. 14. Calculated torsion line after filtering.

A separate MATLAB script reads in the necessary data from the current folder. Equivalently to the bending test, the TL is not used over the whole length. The first and last $50\,\mathrm{mm}$ are not considered for the computation. Again the torsion angle θ is just present as derivative, which would also lead to an unstable result of TS due to the non-perfect dataset. The same procedure as for the BS determination is used. A virtual TL is fitted to the measured one with *lsqnonlin(...)*. The objective function is the relation of the torsion angle, respectively its first derivative θ, the TS and the applied torque M_t. The stiffness value is optimized. Within the objective function, the BVP for the torsion equation is solved by *bvp5c(...)*. In the case of the torsional beam, the formulation is just a ODE of first order. Which means that the relation can be directly used for the solver.

Firstly the problem is solved for a constant TS to generate a good initial value. A lower and upper boundary is again defined in a way, that the given range for sure includes all possible stiffness values. With the appropriate initial value, the problem is solved for the variable TS. The number of SP is adapted to the measurement length with the rule of thumb of one SP per $100\,\mathrm{mm}$. With the resulting distribution of TS and the measured torque, the TL is calculated and compared to the measured one. The error can be investigated to infer the quality of the computation, respectively the measurement.

IV. Results & Discussion

The results for the computed distribution of BS and TS are provided for a test ski. To prove the results, some verification procedures are presented. Considering the design of the STB, the mechanisms are all working properly as well as the sensors. The only sensor which needs a verification is the force sensor used for the torque measurement. Since a mechanism is in between transducer and the twisted ski, a torque measurement with a precise torque spanner is used to prove the accuracy.

A. Bending Stiffness

Since the approach to determine the distribution of the BS is based on the classic Euler-Bernoulli beam theory several assumptions have to be valid. It is obviously not possible to prove a flat and perpendicular cross-section during bending, but it is possible to prove the systems linearity. A linearity measurement is carried out, results are shown in Fig. 15. Measurement points are interpolated with a trend-line to show the divergence of linearity, which is negligible small.

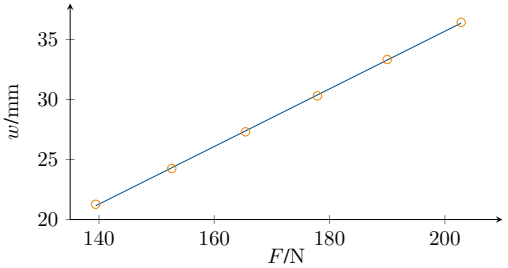


Fig. 15. Linearity measurement for bending test. The ski is bended stepwise and the applied force is measured. The measured points are represented by the bullets. The line represents the linear trend of the measurements.

To verify the BS measurement, the test skis are measured with a three-point bending test. To get an idea of the distribution, 9 measurements are carried out along the ski with a grid of $150\,\mathrm{mm}$. The supports of the bending test provide a distance of $300\,\mathrm{mm}$, where the force is applied in the centre. This method averages the BS over the measurement segments, but provides a good reference and scaling. Outcome is shown in Fig. 16.

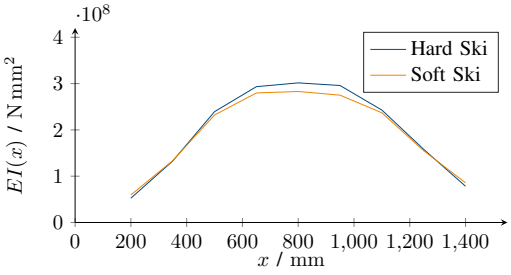


Fig. 16. Three point bending test of both test skis. Measurement points are connected linear to provide an idea of the BS distribution.

Figure 17 shows a determined distribution of BS. In this case for the harder one of both test skis and with 16 SP, meaning that the ski is discretized is segments smaller than

100 mm. By looking at the general shape of the BS distribution, it is quite similar to the one measured with the conventional bending test. The values at the beginning and at the end are quite similar. This decrease obviously makes sense, because the skis thickness is way lower at the front and rear in comparison to its middle section. The thickness has a big influence to the BS, since it is included in the basic formulation with a cubic term.

By investigation of the maximum value, the determined BS provides a higher peak. This also makes sense, because the conventional bending test averages the BS over the measurement region. However, when one is considering the middle section of the ski, meaning at $x = 800\,\mathrm{mm} \pm 150\,\mathrm{mm}$, the mean value of the determined BS will be higher by approximately $10\,\%$ with respect to its maximum value. This higher values and the overshoots could probably be linked to jumps in the layering. For example an aluminium-plate, which is in some ski models incorporated beneath the binding. Such a sudden change of layering, especially at the walling, can highly probably not be measured and determined. This could maybe lead to something like the Gibbs phenomenon. Which means that the overshoot occurs in the region of the sudden stiffness change an can not be described mathematically.

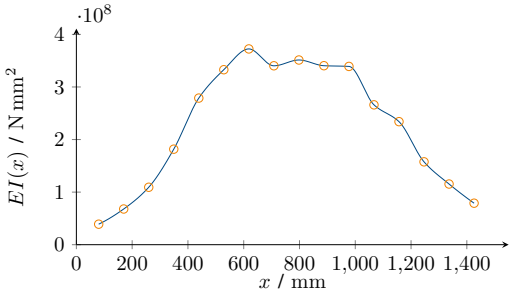


Fig. 17. Exemplary distribution of BS of an alpine ski. Circles represent the SP. Measured model is the test ski with higher stiffness.

The quality of the optimization procedure can be determined by investigating the error e of the BL-fitting. On the one hand the root mean square (RMS) is investigated to illustrate the overall quality. For the given determination, this error is lower than $0.02\,\mathrm{mm}$. Looking at the distribution of the error, local problems can be visualized. Those local errors can be linked to some remaining noise, measurement inaccuracy and other, still not discovered problems.

B. Torsional Stiffness

To prove the concept of the TS determination procedure, again the linearity of the system is analyzed. As shown in Figure 18, a ski under torsion provides a linear relation regarding the applied torque M_t and the torsion angle θ. The measurement is carried out at the STB and over the assumed possible range of elastic torsion, which is in this case $25°$. The measurement points are again compared to a fitted linear polynomial. They do not show a sign of non-linearity. For the torsion test of skis it is unfortunately not possible to make a reference measurement with a standard device. It is not possible to have reference values for the TS.

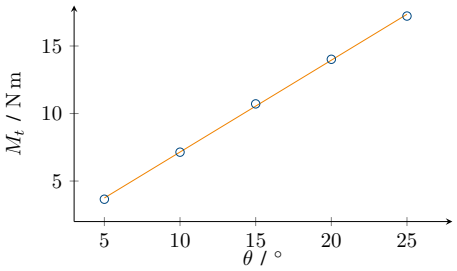


Fig. 18. Linearity measurement for torsion test.

For a validation of the procedure, a test-structure is measured, consisting of a ski-running-surface, GRP, aluminium layer, a steel edge and other polymers. It provides a constant TS over its length with approximately $200\,000\,\mathrm{N\,mm^2}$. This value was determined in another project with an experimental modal analysis (EMA) and with FEM, because a theoretical calculation is not possible due to the composite structure. Figure 19 shows the determined TS for the test-structure. Since this measurement is of a constant distribution of TS, it also the result is quite flat. While comparing the slightly fluctuating determined TS with its mean value, the maximum divergence is within $\pm 1.3\,\%$, which is a satisfying result.

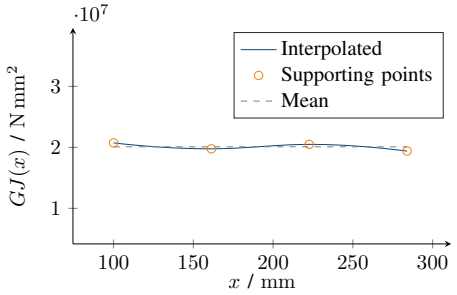


Fig. 19. TS measurement of a test-structure.

When investigating the distribution of TS for the harder test ski in Figure 20, it can be seen that one peak is strongly pronounced. Though, the maximum is in a credible range, the sudden change of the TS does not seem to be realistic. However, the x-position of the peak is in the region with the highest stiffness, because of the highest ski thickness. Nevertheless the uncertainty of this evaluation can probably be linked to the measurement procedure.

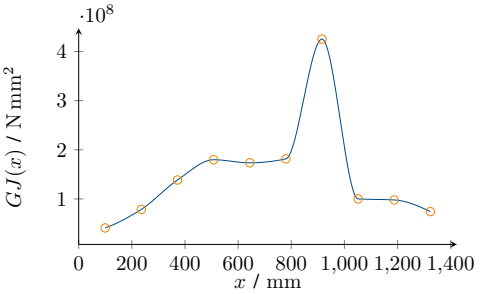


Fig. 20. Exemplary distribution of TS of an alpine ski with 10 SP.

A conventional ski side-shape is shown in Figure 21. As initially planned, the ski is suspended in a way, that the rotation axis is directly at the cross-section. With closer investigations, this is most likely not the best way to apply a torque. Because of the pretension, the rotation axis is by far not on the skis centre-line. As illustrated, the torque M_t rotates the ski properly in the application point of the torque, but not in the centre of the ski. This effect adds an error to the system and thus, the measured rotation angle does not clearly represent the TS of the ski.

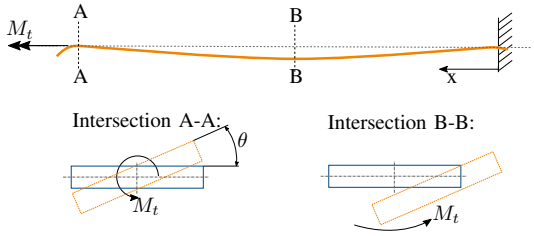


Fig. 21. Pretension of ski leads to deflection of ski under torsion. Rotation and deflection is visualized at different x-positions. Blue represents a flat and even cross-section and orange a ski under torsion.

Beside the problems with the mechanical measurement, the evaluation of the measured data is working properly. The RMS error is approximately $0.02\,\mathrm{mm}$. Fluctuations are slightly higher.

V. Conclusion & Prospect

The determination of bending and torsional stiffness of alpine skis with an indirect measurement procedure is possible and delivers convincing results. Though, the complexity while switching from theoretical calculations to a practical determination is non-negligible. Especially the ansatz, used to solve for the stiffness values varies a lot. While translating the real mechanics with measurements of bending line and applied force as well as torsion line and applied torque into data-sets, small details are having big influence to the determination procedure and results. Starting with measurement precision, a wave-length filtering-procedure and most importantly the data evaluation.

A simple application of the mechanical laws behind the bending and torsional beam is not possible. The occurring derivatives are amplifying measurement uncertainties, measuring object errors and noise to unstable and unverifiable results. However, using a curve-fitting approach comprising the basic mechanical laws leads to reliable results. The generation of a virtual bending line and torsion line, adapted to fit on the measured data provides mechanically verified results. Still, this approach needs more attention considering stability with a different number of supporting points.

Depending on the future application field of the ski-test-bench and the determination of the stiffness values of alpine skis, the mechanism and evaluation of this prototype can be tuned. Meaning that the validation of FEM-models will need a more precise result of the stiffness values. Thinking of a pure comparison between different skis, it is probably sufficient to modify the procedure to a less sophisticated methodology.

From a structural perspective, the STB is working fine. So far, there are no parts which tend to wear out or do not show enough mechanical strength. In considerations of the measurement procedure itself, the measurement head is probably a weak point of the device. It introduces a mechanical uncertainty. A different approach to record the bending and torsion line could maybe lead to a higher accuracy. Also the torque loading of the ski needs to be improved. The mechanism itself is working fine, but the placement of the ski and the mismatch of the longitudinal ski-axis and the torsion-axis introduces some issues.

A more detailed validation of the measurement and determination approach would lead to further insights of the ski-test-bench's reliability. A measurement object with variable but known stiffness values could be useful to check the working principles.

References

[1] Oesterreich.com. Wintersport. [Online]. Available: https://www.oesterreich.com/de/sport/freizeitsport/wintersport

[2] J. Howe, *Skiing Mechanics, First Edition*. Colorado, 1963.

[3] W. Nachbauer, F. Rainer, K. Schindelwig, P. Kaps, *Effects of ski stiffness on ski performance*. 5th Conference and Exhibition on Engineering in Sport, 2004.

[4] J. Truong, C. Brousseau, A. Lussier Desbiens, *A Method for Measuring the Bending and Torsional Stiffness Distributions of Alpine Skis*. Elsevier, 2016.

[5] A. L. Desbiens and M. Bulota. SoothSKI. [Online]. Available: https://soothski.com/technology-measure-ski-properties/

[6] F. Wikerman, *Characterisation of alpine skis*. KTH ROYAL INSTITUTE OF TECHNOLOGY SCHOOL OF ENGINEERING SCIENCES, 2016.

[7] Head Austria Website. [Online]. Available: https://www.head.com/de_AT/ski.html

[8] F.-J. Falkner, *Technische Mechanik 2, Vorlesungsunterlagen*. Department fr Mechatronik, MCI Management Center Innsbruck, 2017.

[9] E. Oate, *Structural Analysis with the Finite Element Method Linear Statics: Volume 2. Beams, Plates and Shells*. Springer Netherlands, 2013.

[10] L. P. Kollar and A. Pluzsik, *Bending and torsion of composite beams (torsional-warping shear deformation theory)*. Sage, 2012.

[11] Matlab Documentation : lsqnonlin. [Online]. Available: https://www.mathworks.com/help/optim/ug/lsqnonlin.html

[12] A. R. Conn, N. I. M. Gould, and P. L. Toint, *Trust-region methods*. Society for Industrial and Applied Mathematics, 1987.

[13] Matlab Documentation : 5vp5c. [Online]. Available: https://www.mathworks.com/help/matlab/ref/bvp5c.html

[14] L. Collatz, *The Numerical Treatment of Differential Equations*. Springer-Verlag Berlin Heidelberg, 1960.

[15] C. Grossmann, H.-G. Roos, and M. Stynes, *Numerical Treatment of Partial Differential Equations: Translated and revised by Martin Stynes*. Springer-Verlag Berlin Heidelberg, 2007.

[16] Renishaw. Tonic Abtastkopf Datenblatt. [Online]. Available: https://docplayer.org/61827526-Tonic-mess-system-datenblatt-l-b.html

Hannes Alexander Schwendinger is a mechatronics master student at the MCI Management Center Innsbruck and emphasises on mechanical engineering. He received a Bachelor of Science in engineering at the MCI at the department of mechatronics before starting his masters programme. Right before his studies, he gained a Masters Craftsman Certificate for Mechatronics and Manufacturing technology at the Austrian Federal Economic Chamber.

Fatigue life simulation process for cold formed parts

Raphael Sigloch and Franz-Josef Falkner (supervisor)

***Abstract*—In many industrial sectors cold formed parts are used. Due to this manufacturing process material behavior changes, which leads also to a change of fatigue life behavior. To be able to make realistic fatigue life predictions on cold formed parts, a simulation process shall be developed, which takes into account material property changes caused by the manufacturing process.**

The goal is to develop a simulation procedure for fatigue life predictions of cold formed parts using Ansys and FEMFAT. Therefore the leading question will be, how can the manufacturing process of cold forming be regarded in a fatigue life simulation procedure.

In the first place a literature research is done with a comparison of estimation methods for cyclic material properties. These are afterwards implemented in Matlab to serve as input for a developed procedure regarding the deformation of the cold forming process. Afterwards a minimalistic fatigue test is performed in reality and simulated with the developed procedure. Additionally the procedure is applied to a lawn mower axis of the company STIHL.

The comparison of the simulation methods shows that the developed simulation process is working and points out differences between the estimation methods in both simulation examples. The performed test does not satisfy the requirements to validate the simulation due to given manufacturing and performing circumstances.

***Index Terms*—Fatigue life behavior, FEA, cold forming, Ansys, FEMFAT.**

I. Introduction

Fatigue life behavior is an essential characteristic to consider when developing any kinds of mechanical components. Unfortunately it is very time consuming and therefor costly for companies to perform fatigue tests on all of their parts. To be able to make a statement about fatigue life in an early development stage, simulation softwares as for example Ansys and FEMFAT exist. With theses software tools the life span of a component under a given load can be predicted. The problem is, that the components undergo a manufacturing process, that can influence the life time, which is not taken into account in theses types of simulations. One frequently used manufacturing processes is cold forming. While cold forming is a widely used manufacturing process in several technical fields, it is not commonly regarded yet in the fatigue life simulations. Due to this reason a simulation procedure taking into account the plastic strains and the according change of material properties caused by a cold forming process is developed in this master thesis.

To achieve this goal previous work dealing with the influence of a cold forming process on cyclic material properties are included in the simulation process. The simulation itself happens in Ansys Workbench with the add-on FEMFAT, performing the fatigue simulation. The calculations for the effect of the cold forming process on cyclic material properties is implemented in Matlab to use them afterwards in the simulation in FMEFAT.

After developing the simulation process a minimalistic test setup is designed to validate the results given by the simulation. The developed method is also applied to a real component, a lawn mower axis of the company STIHL.

In the end the results of the different methods for

R. Sigloch studies at MCI, Innsbruck, Austria, e-mail: raphael.sigloch@web.de

F.-J. Falkner is with the Department of Mechatronics, MCI, Innsbruck, Austria.

handling the process influence are compared to each other and to the results of the test performed in reality.

II. Theoretical background

A. Cyclic material properties

In order to perform a fatigue life analysis it is crucial to know about cyclic material behavior and properties. In the case of this thesis there are three critical characteristics decisive for a fatigue life analysis, the strain-life-curve, the cyclic stress-strain-curve and the S-N-curve. These are described in the following in detail.

1) Strain-life-curve: The so called strain-life-curve consists of an elastic and plastic part as shown in Fig. 1 with slopes b and c. Depicted in a logarithmic scale these two curves end up being straight lines and together forming the strain-life-curve. On the x-axis the load cycles can be seen, while on the y-axis the strain amplitude is given, usually in percentage.

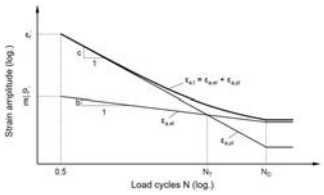


Fig. 1: Strain-life-curve [1]

Manson and Coffin used following equation 1 to describe the strain-life-curve with its exponents b and c defining the slope of each part.

$$\begin{aligned}\varepsilon_{a,ges} &= \varepsilon_{a,el} + \varepsilon_{a,pl} \\ &= \frac{\sigma_f'}{E} * (2N)^b + \varepsilon_f' * (2N)^c\end{aligned} \tag{1}$$

2) Cyclic stress-strain-curve: The cyclic stress-strain-curve looks quite similar to the static stress-strain-diagram, but on its axes amplitudes are plotted. Again it is derived from an elastic and a plastic part, which show as straight lines in logarithmic scaling. The cyclic stress-strain curve is described usually with equation 2 derived by Ramberg and Osgood.

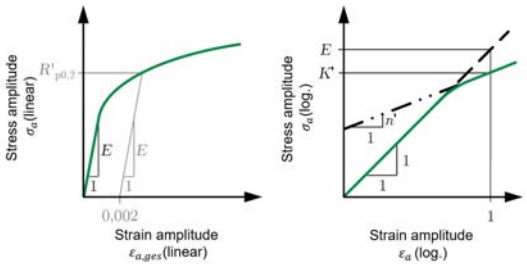


Fig. 2: Cyclic stress-strain-curve

$$\varepsilon_{a,ges} = \varepsilon_{a,el} + \varepsilon_{a,pl} = \frac{\sigma_a}{E} + (\frac{\sigma_a}{K'})^{\frac{1}{n'}} \tag{2}$$

3) Stress-life-curve: The Stress-life-curve, also known as S-N-curve or Wöhler-curve can can be divided in three areas. Low cycle fatigue, finite life or endurance fatigue and durability as shown in 3. It plots the alternating stress amplitude against the number of load cycles N for a material commonly given in a logarithmic scale.

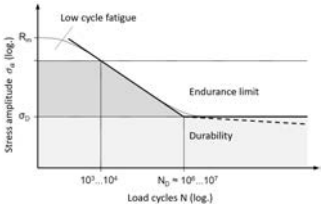


Fig. 3: Stress-life-curve [1]

The Wöhler-curve can be described with equation 3, where k is the slope of the endurance life area. It is also called Wöhlerexponent.

$$N = C * \sigma_a^{-k} \tag{3}$$

B. Fatigue analysis

Fatigue analysis is the structural analysis of failure tendency of a system under cyclic loading. As an outcome one wants to get the life time of a part under cyclic loading. The life time is the time span a material can resist against a specific loading until a crack is initialized or the part breaks.[2]

1) Load data: For a reliable fatigue analysis the occurring loading has to be known in first place. It is given by a load-time history, as shwon in fig. 4.

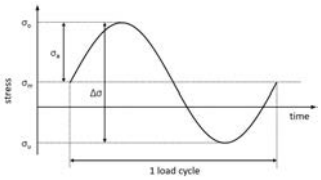


Fig. 4: Load cycle definition [1]

To fully describe a loading it is necessary to know stress amplitude σ_a and mean stress σ_m. As stresses are not always oscillating around zero, the stress ratio R is introduced to define the location of the load cycle. [3]

$$R = \frac{\sigma_l}{\sigma_u} \tag{4}$$

C. Rainflow counting method

The rainflow counting method is a classification method to count occurring load events in a load-time history. It is based on the incorporated energy into the material. The goal is to count closed hysteresis loops, while saving open hysteresis loops as residuals. As output one gets the amplitude and mean of each hysteresis loop and the number of cycles with equal loops. The parameters are depicted in a so called rainflow matrix.[4], [5]

D. Haigh diagram

The influence of the mean stress in FEMFAT is implemented with the help of the Haigh diagram, shown in figure 5, where the tolerable stress amplitude is plotted against the mean stress. It can constructed by the material properties $R_m, R_{p0,2}, \sigma_D$, and gives the tolerable stresses for different stress ratios. [1], [6]

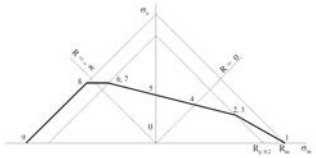


Fig. 5: Construction of the Haigh diagram [6]

1) Influence factors: FEMFAT uses several influence factors in its simulation method to generate a local Wöhler-curve for every node. For this, the material Wöhlercurve is taken and adjusted regarding to local geometry and stress characteristics. The following influence factors are used by FEMFAT for the fatigue analysis. Depending on the user´s will they can be considered or not. [6]

TABLE I: Influence factors [6]

Influence	Fatigue limit	Slope	Cycle limit
Stress gradient	$f_{GR,D}$	$f_{GR,N}$	$f_{GR,E}$
Mean stress	$f_{MI,D}$	$f_{MI,N}$	$f_{MI,E}$
Surface roughness	$f_{OR,D}$	$f_{OR,N}$	$f_{OR,E}$
Technological size	$f_{TG,D}$	-	-
Tempering condition	$f_{VZ,D}$	-	-
Surface treatment	$f_{S,D}$	-	-
Temperature	$f_{TE,D}$	-	-
Statistics	$f_{ST,D}$	-	-
Forging	$f_{TF,D}$	$f_{TF,N}$	$f_{TF,E}$

The above mentioned influence factors modify three particular values defining the stress-life-curve. These are the fatigue limit, the slope and cycle limit.

2) Damage accumulation: In a fatigue analysis the damage on a material or component is calculated according to the so called Miner-rule. The method of miner states that every load cycle above a specific stress level causes some damage on the material. The total damage then is calculated by taking the sum of all partial damages as in equation 5.

$$D = \sum_{i=1} D_i \ = \sum_{i=1} \frac{n_i}{N_i} \tag{5}$$

If the cumulated damage exceeds the critical value of $D_i = 1$ the component fails.
In FEMFAT it can be chosen between three different Miner-rules, differing in the treatment of the high cycle fatigue area. These three methods can be seen in fig. 6.

- Miner Original:
  Stresses under the fatigue limit don´t cause any damage and therefor are neglected
- Miner Modified:
  The method derived by Haibach uses a slope of $2k - 1$ in the area after the cycle limit
- Miner Elementar:
  This method extends the Wöhlerline with its slope k until the zero crossing

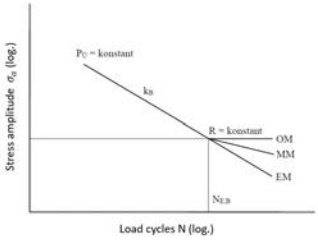


Fig. 6: Miner-rules [6]

III. Estimation of cyclic material properties

There already exists a big variety of studies dealing with the influence of forming processes on material behavior. Three of these methods take into account the equivalent plastic strain φ_v, which can be derived from an forming simulation and therefor will be used in this master thesis. The equivalent plastic strain is calculated as follows:

$$\varphi_v = \frac{2}{3} * \sqrt{\varphi_1^2 + \varphi_2^2 + \varphi_2^2} \tag{6}$$

In 2004 A. Hatscher derived an alternative formulation for the Manson-Coffin-equation for universal supporting points. For this reason equation 7 is used as equation for the strain-life-curves.[7]

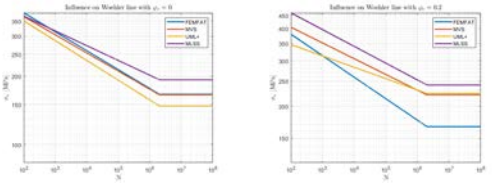
$$\varepsilon_{a,ges} = \frac{\sigma_0}{E} * (\frac{N}{N_{0\sigma}})^b + \varepsilon_{p0} * (\frac{N}{N_{0\varepsilon p}})^c \tag{7}$$

The three estimation methods in table II are regarding the equivalent plastic starin φ_v and are therefor used in the simulation process. They mostly differ in the estimation of the supporting points needed for the description of the strain-life-curve. In [8] a detailed overview of the estimation methods can be found.

TABLE II: Estimation methods for cyclic material properties

Method	Year	Reference
Material Law of Steel Sheets	2000	[9]
Method of variable Slopes 2006	2006	[10]
Uniform Material Law +	2006	[10]

Fig. 7 shows a comparison of the stress-life-curves resulting from the methods with an equivalent plastic strain of 0 and of 0,2. The common cyclic material properties calculated by FEMFAT serve as a reference in blue.



(a) (b)

Fig. 7: Method comparison with $\varphi_v = 0$ (a) and $\varphi_v = 0,2$ (b)

In all three cases it is visible, that they estimate the fatigue limit much higher with an equivalent plastic strain of $0,2$. What is striking, is that the UML+ expects the endurance limit beneath the standard FEMFAT calculation when there is no deformation, while MLSS expects it in this case much higher and MVS is more or less on the same level. Furthermore the influence of the slope can be seen in the right plot, where the UML+ crosses the reference in the low

cycle fatigue region due to its high Wöhlerexponent k.
From the plots it can be withdrawn, that all three estimation methods have a positive effect on the fatigue limit when there is a specific deformation reached. Also it shows that the Uniform Material Law+ is the most conservative estimation method for low deformation. Opposite applies for the Material Law for Steel Sheets.

IV. Simulation procedure

The workflow for the simulation of cold form parts follows the diagram depicted below in fig. 8. The green colored steps are done in Ansys, the blue ones in Matlab and the red ones in Ansys FEMFAT. The white fields are files and data that are handled between the steps.

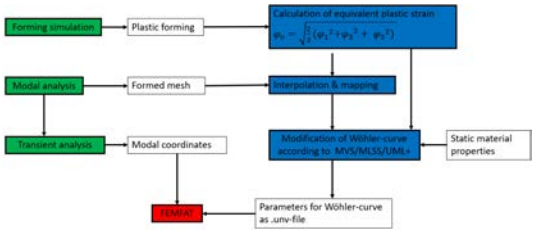


Fig. 8: Workflow of the simulation

Basically it can be divided into three parts executed in Ansys, Matlab and FEMFAT. In Ansys the forming process, modal analysis and transient analysis are performed. For the fatigue simulation some calculations for the forming influence have to be done in Matlab and then given back to Ansys FEMFAT, where the actual fatigue life simulation finally is performed.

The detailed procedure for the simulation is described in the following step after step.

1) *Forming simulation:*
   In the first step the forming process has to be simulated in Ansys with a static structural analysis to get the plastic strain that will be used for the fatigue simulation. With the help of APDL code, the plastic strain is written into a .txt-file.
2) *Meshing of formed part:*
   This step is just to get the coordinates of the mesh of the already formed part. Therefor the mesh is generated inside the modal analysis and again with APDL code nodal coordinates are written into a .txt-file.
3) *Calculation of equivalent plastic strains φ_v*
   Next the in the first step generated plastic strain file is loaded into Matlab and then the equivalent plastic strains φ_v is calculated according to 6.
4) *Mapping of plastic strains on formed mesh:*
   Subsequent the calculated φ_v is interpolated to be able to map them on a new mesh. This is done using the Matlab command *scatteredInterpolant* [11].
5) *Calculation of cyclic material properties:*
   As soon the plastic strain at every node is known, the cyclic material parameters are calculated as described in III inside the according Matlab function.
6) *Generating .unv-file as FEMFAT input:*
   FEMFAT needs the local material properties for every node written in a .unv-file. This is done in Matlab, where the specific parameters are defined given by [12]. In the case of this analysis the tensile strength, tensile compressive alternating stress limit, S-N-curve slope and S-N-curve endurance cycle limit are handed to FEMFAT.
7) *Get modal coordinates:*
   In order to save computational time a mode superposition method is used, which means modal coordinates are needed as input for FEMFAT ChannelMAX. Using the APDL command *TRNOPT, Msup, 3, , , 1*[13] modal coordinates are exported. To make it usable for FEMFAT after the file header *"END_OF_HEADER"* has to be inserted manually and saved as .txt-file.

8) *Define material data:*
   As first step in FEMFAT a material has to be generated. This is done by FEMFAT according to the FKM guideline based on the ultimate strength, that has to be assigned.
9) *Define load history:*
   Next the load history has to be defined with the generated .txt-file containing the modal coordinates from the transient analysis. According to the file format modes and belonging columns are defined.
10) *Insert jobfile command for local material properties:*
    As there is no option to activate local material properties in structure tree of Ansys, a jobfile command has to be added to achieve this.

```
setValue {} {} FatParamMaterial
{2 - Generated Material}
setValue {} {} LocalMatPropRead
{"path and name of .unv-file"}
setValue {} {}
InfluenceLocalMaterialProperties 1
```

11) *Solution:*
    After all the steps before the wished solution output has to be chosen and the simulation can be run finally.

V. Simulation and test setup

In order to make a statement about the reliability of the simulation a fatigue test is done in reality on a minimalist example with a vibrational shaker. Furthermore the simulation procedure is applied to a lawn mower axis of the company STIHL.

A. Test setup

In fig. 9 the test setup for the shaker can be seen. It is consists of an adapter plate which is mounted on the shaker, the specimen and a block to fix the specimen on the adapter plate. The specimen is weakened with a hole at the upper bending position. This is to guarantee the highest stresses and the failure in the formed region, as it is the goal to show the influence of a cold forming process on the component.

Fig. 9: Test setup

The base plate is accelerated on a vibrational shaker near the eigenfrequency of the system to cause enough damage. Therefor in advance of every test a system identification run is performed. This gives the eigenfrequency and damping of the system. Due to varying eigenfrequencies between the tests excitation frequencies differ to make sure not to break the specimen immediately. The acceleration input can only be controlled by the shaker voltage. This results in different acceleration amplitudes for tests, that are measured by an accelerometer attached to the base plate of the test setup.

B. Test simulation setup

Fig. 10: Simulation test setup

For the simulation only half of the model is designed to save computational effort. This is possible due to the symmetry of the test setup. It is also sufficient to only simulate the specimen itself, as adapter plate and fixing block are assumed to be rigid bodies. With these assumptions the bottom face

of the specimen can be accelerated in y-direction with a base excitation as it is shown in fig. 10. The accelerations applied come from the measured acceleration during the tests on the vibrational shaker. All tests are simulated with the three estimation methods by the use of the different .unv-file with cyclic material properties.

C. Axis simulation setup

Additionally the developed simulation procedure is used on the lawn mower axis of the company STIHL depicted in fig. 11.

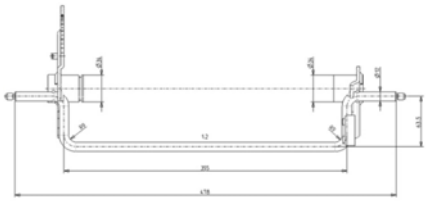


Fig. 11: Drawing of lawn mower axis

For the simulation only one bending region was taken, as from this a forming simulation was available. Also the model was cut in half to save computational effort. The axis than was accelerated by a random signal with an maximum amplitude of $300\,\frac{\mathrm{m}}{\mathrm{s}^2}$. Additionally a load of $300\,\mathrm{N}$ was applied on the face marked in red in fig. 12.

Fig. 12: Drawing of lawn mower axis

The simulation is performed with all three estimation methods and compared afterwards. This is done by changing the .unv-file in the jobfile command.

VI. Results

A. Test

Table III shows the eigenfrequencies of the tests executed. It can be seen there is a discrepancy of $3{,}73\,\mathrm{Hz}$. Together with the low damping of around $0{,}5\,\%$ this makes a huge difference in the outcome of the test. This means with the same shaker voltage input, resulting acceleration amplitudes can vary widely. In order to compare it to a simulation it is therefor necessary to adapt the acceleration in the simulation according to the tests. Test 1 and 2 fo example are excited in the same frequency, but have a life time difference of more than 200000 cycles due to the different eigenfrequencies. For this reason it does not make sense to compare the tests to each other in this. Nevertheless the specimen failed in all tests in the forming region, as it was wanted by inserting the hole.

TABLE III: Comparison eigenfrequencies

	V1	V2	V3	V4	Mean
eigenfrequency [Hz]	$116,25$	$118,67$	$116,06$	$114,94$	$116,48$

B. Test simulation

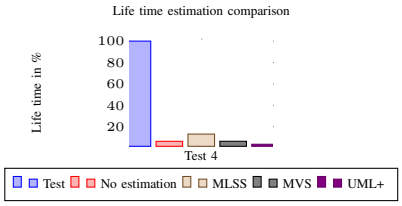
Table IV depicts the results from the simulation of test 4 with an excitation frequency of $112\,\mathrm{Hz}$ and a resulting amplitude of $180\,\frac{\mathrm{m}}{\mathrm{s}^2}$. The results reflect the findings of the method comparison, that has been done before. The Uniform Material Law+ expects the highest damage after 40 seconds and therefor shortest life time. The MLSS on the other hand gives the highest life span for the specimen, while the Method of Variable slopes and standard FEMFAT method lie in between. In all cases the location of the critical node is on the inside of the bending region, except for the UML+, where it is located on the outside.

TABLE IV: Test 4 simulation results

	Node number	Damage after 40s	life time [s]	Tolerable cycles
no method	20801	0,21627	185	21085
MLSS	21233	0,10513	380	43375
MVS	21233	0,21431	187	21278
UML+	20604	0,37936	105	12020

C. Comparison test - simulation

The bar chart below shows the expected life time of test 4 with all estimation methods in comparison with the actual life time measured during the test. It is obvious that none of the simulations comes close to the actual life time and are not even reaching 20 %. The reason for this is, that the test could not be acceleration amplitude controlled. Due to the effect of resonance near the eigenfrequency the acceleration amplitudes are magnified enormously in the simulation. This results in much higher stresses and following this larger damage.
Unfortunately the results from the test cannot be compared reasonably to the outcome of the simulation.



D. Axis simulation results

Table V shows the results from the axis simulation. It can be seen, that the MLSS and MVS expect the axis to withstand the loading longer than with the standard FEMFAT simulation. The Uniform Material Law+ predicts a shorter life time. It is also visible that the UML+ and the standard simulation expect the highest damage in the same critical node. This shows the conservative character of the Uniform Material Law+, as already discussed before. Also MLSS and MVS calculate the same critical node, showing the differences between those two methods. A closer look on the resulting S-N-curves of the critical nodes showed that only the endurance limits differ. This explains why the effect of the expected fatigue limit can directly be seen in the damage results.

TABLE V: Results axis simulation

	FEMFAT	MLSS	MVS	UML+
Critical node	26143	26182	26182	26143
Damage (40s)	$3,12e^{-4}$	$1,26e^{-4}$	$2,69e^{-4}$	$5,84e^{-4}$
Life time [h]	35,6	87,95	41,38	19,03

VII. Conclusion & Outlook

As a conclusion from this master thesis it can be said, that a working simulation procedure, that regards the influence of cold forming has been developed. The methods found in the literature research have been implemented using Matlab so they can be used for a fatigue life analysis with FEMFAT inside Ansys. Also the mapping of equivalent plastic strains given from a forming simulation onto the mesh for the fatigue life simulation can be done inside Matlab. Of course the accuracy and correctness of the mapping process is strongly dependent on the accordance between the geometry after a forming simulation and the geometry used for the fatigue life simulation. Furthermore it is clear to say that the fatigue simulation regarding the influence of a forming process can only be as good as the forming simulation is. Additionally, the more is known about used material the more accurate the simulation results will be.
From the simulations of the test and the lawn mower axis it could be shown that the simulation procedure is working. Also differences between the estimation methods and their behavior under given deformations were pointed out.
Unfortunately the test performed to validate the simulation could not do so. This was due to the fact that the vibrational shaker could not be amplitude controlled. This together with the low damping of steel resulted in a huge discrepancy already between tests caused by the excitation near the

natural frequency of the system. In the simulation this amplification resulted in huge damage values underestimating the life time of the specimen by far.

As a future outlook it is recommended to perform a test on a vibrational shaker, that can be excited at least amplitude controlled. In an ideal case the test can be executed strain controlled. This would give the possibility to directly compare the strain-life-curves from the estimation methods, as they are based on it. It is also suggested to have a specimen with a tapered shape in the forming region similar to the shape of a quasistatic tensile test. This reduces the influence of the notch factor and therefor high stress gradients as it was the case in the tests performed.
Furthermore it would make sense to test a real component, for example the STIHL lawn mower axis used for the simulation. Of course it than has to be guaranteed to ensure a failure in the forming region to see the effect of the cold forming influence on the fatigue life behavior.

ACKNOWLEDGMENT

First of all I want to thank my supervisor of the MCI, Dr. techn. Franz-Josef Falkner for his guidance and support throughout this thesis. Also I want to thank Matthias Panny for his help during the tests for the thesis. Furthermore I want to thank the company STIHL, in person of Dr. Manfred Hofer, for providing data and meaningful discussions.
A special thanks goes to Sebastian Bauer for providing the input for my simulations with the forming simulations of his master thesis.
Last but not least I want to thank my family, my girlfriend and friends for always supporting me during this time and their patience. Without all of you this would not have been possible.

REFERENCES

[1] S. Götz and K.-G. Eulitz, *Betriebsfestigkeit*. Wiesbaden, Germany: Springer-Verlag, 2020.
[2] (2022) What is piping. [Online]. Available: https://whatispiping.com/basics-fatigue-analysis/
[3] V. Läpple, *Einführung in die Festigkeitslehre*. Wiesbaden, Germany: Vieweg + Teubner, 2011.
[4] Dr. M. Hofer, "Lecture notes in fatigue - loading," October 2020.
[5] M. Köhler, S. Jenne, K. Pötter, and H. Zenner, *Zählverfahren und Lastannahme in der Betriebsfestigkeit*. Heidelberg, Germany: Springer-Verlag, 2012.
[6] *FEMFAT 5.3 Basic Theory Manual*, Magna Powertrain, 2018.
[7] A. Hatscher, "Abschätzung der zyklischen Kennwerte von Stählen," Ph.D. dissertation, Technische Universität Clausthal, 2004.
[8] M. Wächter, "Zur Ermittlung von zyklischen Werkstoffkennwerten und Schädigungsparameterwöhlerlinien," Ph.D. dissertation, Technische Universität Clausthal, Clausthal, Germany, May 2016.
[9] R. Masendorf, "Einfluss der umformung auf die zyklischen werkstoffkennwerte von feinblech," Ph.D. dissertation, Technische Universität Clausthal, Clausthal, Germany, 2000.
[10] A. Hatscher, T. Seeger, and H. Zenner, "Abschätzung von zyklischen werkstoffkennwerten - erweiterung und vergleich bisheriger ansätze," MP Materials Testing, Tech. Rep., 2007.
[11] (2022) Mathworks. [Online]. Available: https://de.mathworks.com/help/matlab/ref/scatteredinterpolant.html
[12] *FEMFAT 5.4 Basic User Manual*, Magna Powertrain, 2020.
[13] Ansys, *Command Reference*, 2022.

Raphael Sigloch is a master student at the MCI, Innsbruck studying "Mechatronics and Smart Technologies" with a major in mechanical engineering.

Potential improvements to the resistance spot welding process in the car body shop by means of company internal benchmarking

Bernhard Spornraft, Barbara Hansen (supervisor) and Benjamin Massow (supervisor)

***Abstract*—In today's car body shop many different sheet metal parts are joined by spot welding operations. Many demands are applied to the welding task when it comes to the safety crash test of the car, mechanical durability of the welding spot, and process stability. These are in place to deliver a high-class end product. The spot welding performance depends on several adjustable parameters and a wide range of rigid circumstances that have an integral influence on product quality and manufacturing costs. For this reason, a steady improvement is required. Yet, this is hard to achieve as the process is already in a well advanced stage. Within the company of interest, the product manufacture takes place in different plants globally. Especially, the present car body shop shows a performance difference in comparison with other plants. How these plants operate performance related procedures regarding the spot welding process is unknown. Within this study the manufacturing process of resistance spot welding in the car body shop is analysed. The organisational structure is clarified including the different departments, their competences and connection to each other. Subsequently, the resistance spot welding process is investigated with respect to its performance indicators and crucial influencing fields only within the car body shop. The results of the process analysis are used to create a benchmarking plan to compare the spot welding process in two additional plants. This allowed certain performance indicators and the handling of defined process influencing fields to be compared. The results are subsequently evaluated. Conclusions towards possible optimisation potentials in the present plant are made. These are ranked in terms of costs, benefits and risk. The conducted investigations revealed that the current handling of the spot welding process has many improvement potentials. The manufacturing process could be optimised in terms of staff capacity and possible equipment configuration that has not yet been established. Furthermore, knowledge exchange between the plants should be increased in the field of teaching welding parameter. In summary, the company-internal benchmarking is a very useful tool to obtain optimisation potentials. The discovered potentials can be implemented with the security of a reasonable investigation of the whole manufacturing process. There are, however, other unidentifiable fields of improvement due to similarities between the compared plants. For this reason external partners should also be considered.**

***Index Terms*—Car body shop, Resistance spot welding, Company-internal benchmarking, Process optimisation**

B. Spornraft is a student at the Department of Mechatronics, MCI, Innsbruck, Austria.

I. Introduction

In the mass production of the automotive industry, several manufacturing steps need to be done in order to provide a reliable, safe and superior end product. One of these steps is within the car body shop where the mechanical framework of the car is produced by joining steel sheets together [1]. The main joining technology in this sector is the resistance spot welding. In this procedure an electrical current is sent through metal sheets and by means of electric resistance the welding heat arises and forms the welding spot [2]. Due to progressive demand, continuous refinement of the resistance spot welding process is absolutely necessary to reinforce efficiency. To improve a process such as spot welding, potential sectors of optimisation need to primarily be identified. However, resistance spot welding is a sophisticated process with wide ranging influencing factors [3]. Identifying potential optimisation measures is not so straightforward.

To identify potential indirect improvements new approaches have to be considered. Between the supported plant and the other nearby plants there is an obvious performance gap in the resistance spot welding process. This is particularly obvious when looking at performance-related indicators recorded during the operation of the manufacturing process. To investigate why other plants show better process execution values than the own plant, a company-internal benchmarking is performed to identify the possible optimisation fields of the present manufacturing process. This also involves understanding the organisational structure and process of the resistance spot welding within the supported plant. These optimisation potentials are then compared with those of partner plants in the same company. The intention of this comparison is to detect any significant differences which are then evaluated and defined. Ultimately, the purpose of this investigation is to optimise the resistance spot welding process towards the performance of other plants. The product quality and the productivity, however, must remain unchanged. Finally, the multiple potentials are rated in terms of risk, possible benefits and required effort.

II. Methods

This chapter illustrates the methodological approaches used to achieve the obtained results. An overview along with the corresponding order of the investigated topics can be seen in Figure 1.

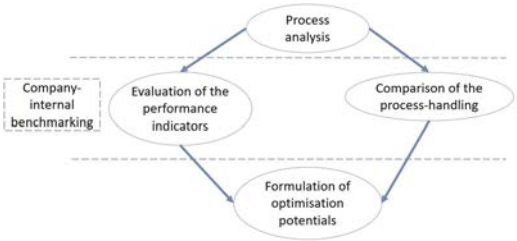


Fig. 1. Overview of the investigated topics

To define specific optimisation potentials for the car body shop of the supported plant, a process analysis followed by a company-internal benchmarking is done. The process analysis identified the most influential areas of the resistance spot welding-process on the car body shop. This included a review of the organisational structures and available performance measures of the resistance spot welding-process. This is then followed by a company-internal benchmarking where the resistance spot welding-process in the car body shop of various

factory plants are compared in terms of performance indicators (performance benchmarking) and process handling (process benchmarking). The results of this analysis allows the formulation of optimisation potentials.

A. Process analysis

Manufacturing a car body follows a defined procedure. There are also many additional requirements, such as safety, that the car body must fulfill [4]. Due to the complex and diverse requirements imposed on a car body, the company established an organisation with many departments, each with their own responsibility, concerning the car body production. Therefore, the first part of the process analysis is to identify the organisational structures within the company and their involvement in the resistance spot welding-process. This consequently gives insight into company-internal guidelines and procedures regarding the resistance spot welding-process. Organisational charts serve as visual aids on how the car body shop is arranged. Furthermore, any available performance indicators of resistance spot welding-related areas are analysed with regards to their significance on the resistance spot welding-process. For the sake of transparency, these indicators are assigned to certain departments of the car body shop. Additionally, elements within the process handling with a great influence on the performance of the resistance spot welding process are determined and their performed procedures are further analysed. The necessary data is gathered with the aid of a questioner.

B. Company-internal benchmarking

In the company-internal benchmarking, different production sites are compared to each other. For this a benchmarking plan had to be designed which followed a defined course based on the approach of *Spendolini* [5]. The design breakdown is as follows:

1) Determine what to benchmark
2) Identify benchmarking partners
3) Collect and analyse benchmarking information
4) Take action

The process analysis served as a basis for point one, as prior knowledge of the resistance spot welding-process is needed in order to define areas to be benchmarked.

1) Performance benchmarking: The fist type of benchmarking applied in this thesis is the performance benchmarking. This type focuses on the performance units of the investigated process and compares them with those of the benchmarking partner. Therefore measurable quantities are of great interest in performance benchmarking. According to [6], meaningful performance indicators have to fulfill certain criteria. In the case of the car body shop these criteria along their corresponding explanation are listed below:

- Topicality: An indicator is topical when it corresponds to the actual circumstances of the considered operational area.
- Acceptance: An indicator is accepted when it is used by management and simple shop floor operators as a source of performance measurement.
- Compatibility: Indicators that can be looked up with given information systems are considered compatible.
- Relevance: An indicator is relevant when it depicts only decision-relevant data. A positive trend of this indicator is in connection to a significant performance improvement.

- Validity: The validity of an indicator means that the used data recording method has the required reliability.
- Comparability: An indicator is comparable if it is equally interpreted by various company-internal institutions. Also a comparability of different time points has to be given.

Some of the in the process analysis defined indicators had to be modified in order to fulfil the described criteria.

2) Process benchmarking: Process benchmarking concentrates on the daily operations of a process. The improvement of day-to-day operations can have a positive impact on the overall process performance [7]. In the case of the resistance spot welding-process in the car body shop, daily tasks are analysed by a process analysis. A few selected tasks are identified as having a great impact on the overall performance of the process. Subsequently, the daily tasks chosen are compared and evaluated using the previously defined benchmarking plan to find differences in the process handling between the factory plants.
Since the chosen tasks are very extensive in their execution possibilities across the plants, any substantial comparative analysis needed a structure. For this reason an expert questionnaire is designed. Based on [8] the creation of such a questionnaire follows a defined procedure.

1) Research question: At first the research question is defined. As mentioned before, the goal of process benchmarking is to find differences in the resistance spot welding-process handling between the investigated car body shop's.
2) Analysis dimensions: The research question is then translated into analysis dimensions. Based on the performed process analysis, the strongest influencing elements of the resistance spot welding-process performance are defined. These elements are later defined as the analysis dimensions.
3) Question complexes: In the next step, the analysis dimensions are further divided into question complexes. This is achieved by splitting the defined influencing areas in several sub-procedures.
4) Interview questions: Finally, the interview questions are formulated from the question complexes. In the course of this, the sub-procedures are further divided into concrete tasks of interest.

Out of the results of the diverse analyses the optimisation potentials are deduced. In particular, the most significant differences in the process handling procedure are associated with performance indicators of the corresponding process-area. Furthermore, cause-and-effect relationships are constructed with performance indicators and specific resistance spot welding-related procedures based on the process analysis. In order to raise the acceptance of the claimed optimisation potentials. Finally, a ranking of the optimisation potentials is created depending on their realisticity, available resources and required working-effort. This is done in order to introduce the subjective but substantiated advice on the direction the 'next step' should take.

III. Results

A. Process analysis

The process analysis clarifies how the car body shop is organised and includes a descrip-

tion of the single departments, their expertise and connection to each other. The corresponding visualisation of the overall structure can be seen in Figure 2.

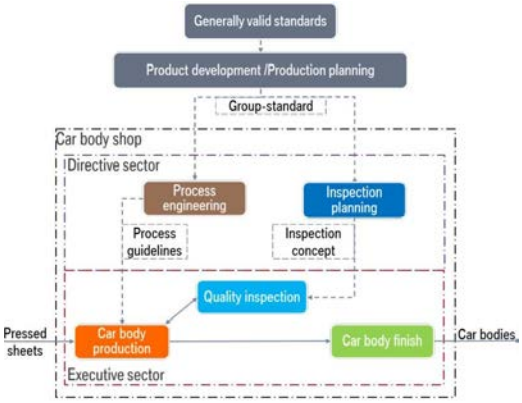


Fig. 2. Overall structure of the car body shop

1) Organisational structure: Basically the car body shop consists of two sectors. The executive sector carries out the required production steps of the car body manufacturing and performers quality inspection measures. The directive sector provides crucial specifications for a successful machine operation and defines the quality inspection procedure. As depicted in 2 this is the connection between the two sectors. Another important department of the car body manufacturing process is the production development and production planning which translates generally valid standards such as norms to guidelines of the directive sector. The company internal designation of these guidelines is the so-called 'Group-standard' which is generally valid in every plant of the factory.

2) Quality capability: Subsequently, the resistance spot welding process in the car body shop of the supported plant is investigated. On this occasion the quality-capability process needed to justify, that a legally demanded product quality can be achieved by the resistance spot welding process [4]. First of all there is the planning phase of a new product where prerequisites regarding the product and manufacturing facilities are determined. With regards to the organisational structure this is performed by the production development and production planning departments. After this the commissioning phase starts where the results of the planning phase are implemented in a real production environment. The aim of this phase is to increase the process and product quality constantly by means of steady optimisation operations. The commissioning phase is further separated in four building stages, the so-called 'pilot production' (PP). At the end of each stage, specified product and process values related to quality and productivity have to be achieved. In Figure 3 it can be seen how the preferable course of the product and process quality looks like. The end of the commissioning phase goes along with the start of production (SOP).

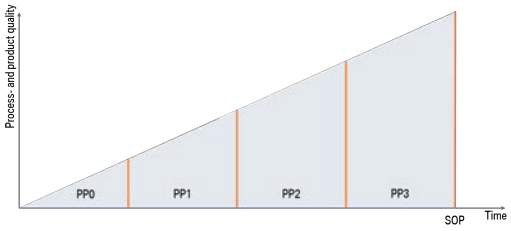


Fig. 3. Building stages in the commissioning phase

The last product phase is the serial production phase, where the car bodies are manufactured within a specified cycle time due to economical

reasons. However, the product and process quality has to be kept at the same level as at SOP. For this reason, constant quality tests are carried out by the quality inspection department. The manufacturing process quality kept constant due to defined maintenance and optimisation measures, carried out by the staff of the car body production department.

3) Resistance spot welding in the car body shop: Performance indicators related to the resistance spot welding process in the car body shop are defined within the process analysis due to an expert questionnaire. The significant performance indicators are assigned to corresponding departments of the executing section and can generally be categorised as either input or output indicators. Input indicators refer to contributed resources, such as number of employees, or amount of quality inspections. Output indicators refer to process-performance-related numbers such as productivity, or achieved product quality. In the case of the car body production department the input indicators are the number of people deployed to maintain the resistance spot welding process in the shop floor (so-called 'process specialists'), the welding control in use and the different sheet combinations applied in the products. The spatter quote describes the number of welding spatter, occurring relative to the overall executed welding operations on one day. It serves as the most important output indicator in the car body production with relation to the welding performance. The second detected output indicator is the welding-error quote which provides information about the occurrence of production stand still due to welding-errors of one day. The input performance indicators of the quality inspection focuses on the resources, provided in order to ensure a proper assurance of the quality of the resistance spot welding process. In this context, the safety concept in the commissioning and serial production phase is looked at (with regards to the investigated product phases). Furthermore, the number of serial and special destroying tests, within a period of half a year, is detected as input indicators since they show the effort performed to assure a certain level of product quality. The output indicator in the quality inspection is the so-called 'failure-quote' which describes the number of appeared diameter violations in the destructive test of welding spots within one half year. The limit of this quote is defined by the product development department. The value of this quote must be under this specified limit because of legal requirements for passenger safety. The failure quote is calculated by dividing the appeared failures by the overall number of tested welding spots. The only detected input indicator related to the resistance spot welding process in the car body finish department is the number of employees in the final inspection. Since the car body shop has to deliver a certain internal quality to the paint shop, final optical inspection exists within the car body finish. If a pollution on the body skin is detected, manual rework measures are introduced. Since the greatest source of pollution are burns from welding spatters, the number of employees needed to ensure the specific internal quality are considered as a measure of the spatter performance.

The last part of the process analysis focuses on important influencing elements of the resistance spot welding process. In particular the teaching process of welding parameters is further investigated. This takes place within a laboratory where welding parameters are created

for every required sheet combination by means of therefore provided test sheets. Welding parameters have a large influence on the welding performance in the production. All welding parameters are provided for material combinations in the commissioning phase of the product. The parameterisation process takes place within a so-called teaching cell where production-related disturbances can not be simulated. For this reason welding parameters can have a great performance in the teaching process but fail in the real production. Consequently, welding parameters have to be optimised in the serial-production phase by the production staff. In order to categorise the parameterisation process the procedure is split up in four main steps. The first step is the communication that a new parameter is necessary in the production. The kind of information transmitted within this stage are of interest. Secondly, the previous welding parameter and other circumstances are analysed. In this context, the adjustable values within a welding parameter are crucial. Furthermore, material related circumstances such as type of material and coatings are investigated. Additionally, geometrical and equipment-related circumstances can be looked at within the analysis. The next step is the teaching process where the new welding parameter is created due to the previously performed analysis. Within this parameter creation process the welding current, welding time and electrode press forces are adjusted. Moreover, special modification such as pause time, current slope and pre-impulse can be implemented. The corresponding parameter creator decides if the created parameter curve can be used in the production according to certain quality characteristics which are recorded during the welding operation. Finally the created parameter has to be secured by means of a destructive test of the welded spots. In this context the diameter of the welding spots are measured after the destruction. Only if the specified minimum diameter is reached by every welding spot, (which is related to certain curves of the parameter set) the parameter is used in the production.

The second influencing element investigated within the process analysis is the procedure of spatter optimisation in the production. In this context it has to be mentioned that an exorbitant spatter quote is one reason for optimisation measures in the commissioning and serial production phase. The first step is the analysis of circumstances of the welding spot to be optimised. On this occasion the equipment, the part geometry and the welding parameters are checked for possible failures. Secondly optimisation measures are performed according to the previous analysis. Generally, equipment-related measures are carried out at first since they cause less effort in process assurance. After that parameter-related optimisation measures are executed. On this occasion it has to be taken into account that the welding energy is decreased in order to reduce welding spatter. For this reason parameter-related optimisation measures have to be performed carefully and needs advanced knowledge about resistance spot welding. The last step of a spatter optimisation is the assurance process. Depending on the type of performed optimisation either destructive or non-destructive assurance methods are applied. This is defined by the process engineering department with regards to the organisational competences.

B. Company-internal benchmarking

After the process analysis is done, the detected performance indicators and influencing elements are compared to other factory plants. In particular the performance values (performance benchmarking) and procedures (process benchmarking) of the present plant (W01) are compared with two other plants designated as W02 and W06 in the following. The reason these plants are selected is because of the close geographical location and the comparability of the used equipment and staff in the car body shop.

1) Performance benchmarking: The performance benchmarking aims to compare the performance indicators of the resistance spot welding process in the car body shop from the process analysis. Here, the most important production and quality-related performance indicators are compared by visualising and interpreting the recorded data. By looking at a number of used welding control (WC) in conjunction and the number of deployed process specialists (PS) (Figure 4), it can be seen that W01 uses the most staff-related resources in order to handle the resistance spot welding process.

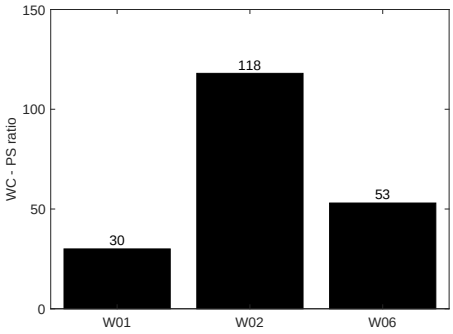


Fig. 4. welding control - process specialist ratio

In this context the number of this ratio means how much welding control one process specialist has to maintain generally. However, W02 and W06 deploying more employees in the creation process of welding parameter than W01. The comparison of the number of teaching specialists demonstrates that there are big differences between the plants. In W06 it is common that the majority of the process specialists (10) are able to teach welding parameters. However, they are usually working in the shop-floor. In W02 there are two process specialists doing the main parameterisation work, both of which are generally deployed in the teaching cell. The third person who is able to teach parameters is the employee of the process engineering-department. In W01 one process specialist and one person of the process engineering-department are teaching the parameter. Both of them also have to fulfill tasks in the production. Also the number of process specialists deployed, to optimise the welding process in the commissioning phase is different. Here W02 (7 people) and W06 (6 people) are at a higher level than W01 (5 people). Further evaluated input indicators show

that W02 (30.1 %) has the highest percentage number of sheet combinations with ultra high strength steel, whereas W01 (17.9 %) and W06 (13.8 %) have similar values. This sort of material seems to be difficult to join using resistance spot welding leading to the assumption that the welding tasks in W02 are generally more difficult. Now the output performance indicators of the resistance spot welding process are shown. At first the evaluated total spatter quotes of the benchmarking-plants are depicted in Figure 5.

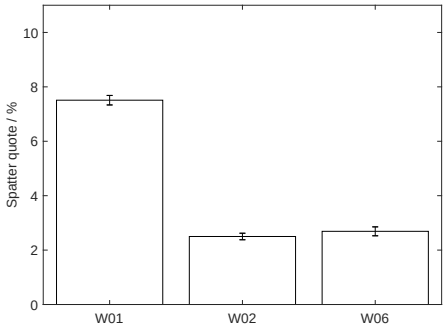


Fig. 5. Total spatter quote in the serial production

The standard deviation shows that the evaluated values do not vary in a large range which indicates that the numbers are trustworthy. It

TABLE I
VALUES OF SPATTER QUOTE AND STANDARD DEVIATION

Plant	W01	W02	W06
Spatter quote / %	7,51	2,50	2,69
Standard deviation / %	0,18	0,12	0,16

can be seen that the spatter quote of the W01 is approximately three times higher than in the other two plants which indicates the urgent need of process optimisation. However, the welding error quote is at the same level conveying that the resistance spot welding process runs at the same productivity level. Next the performance indicators of the quality inspection of each plant are compared as well as the production-related indicator values, first from the input and then from the output perspective.

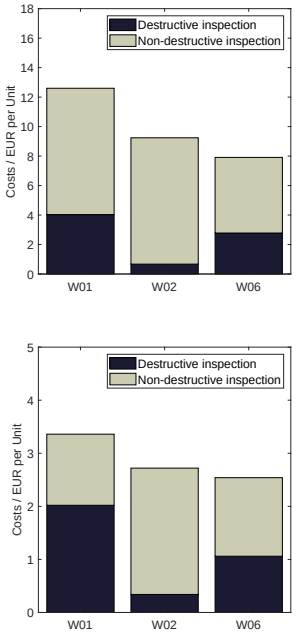


Fig. 6. Costs per unit regarding the inspection concept

This graphic 6 depicts how different the inspection concepts are to each other. In general, W01 uses the most resources in the commissioning phase (left figure) for the required process and product quality. In the serial production phase (right figure), all plants have a similar cost level for the inspection concept. In the commissioning phase W01 and W02 have the

same value of non-destructive test. However, in W01 the share is reduced a lot in the serial production contrary to W02. A further investigation of the numbers of inspection concepts, show that there is also a big difference in the applied test procedures between the plants. Due to the different inspection concepts the output values of the performed cyclical and special inspections within a certain period of time (half year) are very unequal. A best practise can not be formulated in this case but it can be mentioned that W01 has the highest applied effort according to 6. However the destructive failure-quote, recorded for a specific type of car body in the serial production phase, can be compared very good. This is depicted in Figure 7.

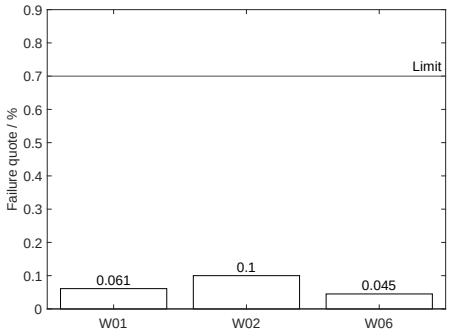


Fig. 7. Failure quote of a certain type of car body in the serial production phase

This quote shows that the quality assurance of W01 detects less diameter valuations than W06. W02 is still far below the specified limit of 0.7%. Due to the risk of the applied destructive procedure and amount of performed tests the concept of W01 can be stated as the most safe.

2) Process benchmarking: In the present case, the process benchmarking concentrates on the differences of the daily operations of a process. The improvement of day-to-day operations can have a positive impact on the overall process performance. [7] In order to evaluate the investigated procedures according to the process analysis characteristics, related to the performance of the teaching process, had to be defined. In the case of the teaching of welding parameter these are:

- Simplicity: Easiness of performing the steps of parameterisation
- Production-reference: Reference of the welding parameter to the production
- Innovation: Innovative thinking is needed in parameter teaching due to increasingly harder circumstances of welding in the production
- Guideline conformity: Since there are some overall guidelines provided by the production planning department, it is important that these are satisfied.

Some of the described characteristics are not compatible for every procedure step. Furthermore, in a few cases one characteristic is contrary to another. The single procedure steps are evaluated according to characteristics. The results of the evaluation show that W02 and W06 are pretty close to each other, even though both have a very different procedure. In general W02 strengths are due to high innovative power and production-reference, whereas W06 possesses good values at the simplicity and production-reference. W01 has the best values in the guideline conformity. Moreover, the simplicity seems to be positive as well.

Next the comparison of the spatter optimisation is inspected. Therefore, the performance-

related characteristics are described in order to evaluate the procedures of the benchmarking plants:

- Safety: The safety of a spatter optimisation concerns itself with the risk that a performed measure can bring quality-related issues. Any actions that increase the risk of producing a lower welding spot quality are rated with a lower safety value.
- Rapidity: The procedure steps are compared to their corresponding rapidity which is very important especially in the serial production.
- Optimisation quality: This point refers to the optimisation quality of the executed procedure steps.

In the evaluation of the different spatter optimisation procedures W02 has the highest total score prompted by the optimisation quality and rapidity thanks to the introduced spatter-optimisation-tool and enabled parameter teaching in the manufacturing cell. W06 reaches a good value in rapidity due to the proportional reduction of the current curve which is an easily administered optimisation measure. W01 convinces mainly in the area of safety because every change to the welding parameter in the manufacturing cell causes a start and end milled destructive test. In doing so, the worst possible cases are usually considered.

C. Optimisation potentials

Finally the optimisation potentials are deduced by the results of the company internal benchmarking. Based on the organisational structure, the potentials are assigned to the executive and directive sectors.

1) Executive optimisation potentials: The first improvement potential is the introduction of a teaching specialists committee for the plants located in Germany. One overall recognition of different procedures of the welding parameter teaching is, that there are many interesting approaches of parameterisation in the investigated plants, even though the teaching process is performed equally. The performance of the welding parameters is very much dependent on the expertise of the corresponding specialist. However, a so-called 'best-practise' parameter set-up out of specific material combinations is not yet known. This is because it is not known how teaching specialists of other plants deal with specific material combinations. For this reason introducing a teaching specialist committee could resolve this issue. Possible functions would include discussing exclusive issues related to the parameterisation process including parameter modifications, preferable structure of a parameter-set and handling of specific material combinations.

The first benefit is the learning value for every teaching specialist regarding the quality of the parameterisation. Since details about the welding parameter creation are discussed within the committee, teaching specialists in turn have to justify their procedure. Out of this the 'best-practise' single material combinations, in terms of teaching the parameter, can be formulated at least internally in the company. On a long-term basis the formulated best-practise suggestions can be written down and used as a great guideline for new teaching specialists. This guideline secures the knowledge regarding teaching process of the company.

The next optimisation potential is to enable and qualify process specialists to optimise the

welding parameter directly in the manufacturing cell. It has to be mentioned that the execution of this optimisation measure requires an enhanced knowledge of the resistance spot welding process and parametrisation. For the sake of error prevention, only experienced process specialists should be allowed to carry out this optimisation procedure. A further error prevention measure would be an on-site controlling of an experienced person to determine, if the optimisation is really necessary or reasonable. The exemplary model is W02 where this procedure is common practice.

The component-specific teaching process has a positive effect on the welding performance of welding spots difficult to optimise. In many cases the welding parameter created by the teaching cell do not fit perfectly. This is because many disturbances present in the production cannot be simulated in the parameterisation process. So, a learning process regarding the connection of welding parameter and disturbances at the manufacturing of the process specialist is appointed to carry out this optimisation measure begins. Furthermore, there is a motivating aspect for the selected process specialist as their work gets more challenging due to increasing complexity of the tasks.

2) Directive optimisation potentials: The formulated potentials are addressed to the process engineering and inspection planning departments. On this occasion, the first potential is to increase the parameter teaching capacity. As conveyed in the results of section performance benchmarking, the number of people who deal with the welding parameter teaching is very different in each investigated plant. However, the potential improvement to performance in the spot welding process is huge in this area. For this reason the number of teaching specialists should be increased. An exemplary model here is W06 where the parameterisation process is carried out by approximately 10 empowered process specialists. The philosophy is to keep the teaching process rather simple, so that many people can be instructed on the know hows. The complexity of the parameter sets in W01 is similar to W06, but there are fewer workers able to perform the teaching process. This has a negative impact on the duration it takes for a new parameter to become available. W06 has more teaching specialists but they are somehow less experienced, so the process engineering department in W06 validates that parameters created by the teaching specialists. An advantage of this is the lowered risk that wrong parameters get into the manufacturing cells.

First of all, it is advantageous that process specialists are as familiar with the resistance spot welding process as possible. This is achieved by increasing the number of process specialists who are able to teach welding parameters. The possibility that the process specialists increase their process-related knowledge is huge. The reason for this is on the one hand, the high influence of welding parameter on the resistance spot welding process and on the other hand, the high configuration options in the parameterisation.
In the long term the pool of experience owned by the process specialists can be the key to a far better process performance.

The next improvement potential is to built up a spatter-optimisation-tool. This potential refers to the analysis procedure in W02 where a certain data-tool bundles all the data, necessary for a spatter optimisation. Since the required data is already recorded in W01, establishing a so-

called 'spatter-optimisation-tool' is just a matter of programming effort. W01 uses the business intelligence tool *Power BI* in order to visualise data recorded every day in the production such as welding spatter quote, welding-error quote or results of the peel tests. Concerning the already existing expertise and availability of data in *Power BI* it would be recommendable to built up the spatter-optimisation-tool within this tool. The first benefit is, with the aid of the spatter-optimisation-tool, the analysis of the data-related circumstances becomes standardised as the only data to be analysed to improve operations. This means, the analysis to determine which optimisation measure needs to be carried out, becomes quicker. In the same way it gets better in terms of the analysis quality where less data-related signs can be overlooked.

3) Ranking of the potentials: Lastly the described potentials are ranked in terms of the effort-benefit ratio. Furthermore, the risk of the potentials to lower the process or product quality are considered. This ranking can be seen in Table II.

TABLE II
RANKING OF THE OPTIMISATION POTENTIALS

Potential	Effort	Benefit	Risk
1. Increasing teaching cap.	-	++	+
2. On-site param. teaching	+	++	–
3. Teaching spec. committee	-	+-	+
4. Spatter-optimisation-tool	+	-	+-

IV. CONCLUSION

The aim of this thesis is to identify optimisation potentials in the resistance spot welding process in the car body shop of the supported plant, by means of a company internal benchmarking. The indicator of the present paper is an obvious performance difference between the resistance spot welding process in the different factory plants located in Bavaria. To close this performance gap potential optimisation areas must be identified. According to the ranked optimisation potentials, the teaching capacity should be increased. For further clarification the training content and number of process specialists to be instructed have to be defined. The required resources have to be provided and the responsibilities have to be divided between the process specialists and process engineering department. Further optimisation potentials could be gathered by performing a benchmarking which included additional factory plants. It had to be evaluated how local circumstances of each plant are compared. Moreover an external benchmarking, regarding the resistance spot welding in the car body shop, would yield many additional approaches of process handling. However, a major obstacle for this procedure is the accessibility of the required data. Finally, new ways of improving the spot-welding process should be considered as well. Quality inspection and parameterisation show very promising new approaches, although this has to be done by the production planning department.

ACKNOWLEDGMENT

I would like to thank the following people for helping with this research project: My MCI supervisor Mr. Benjamin Massow whose insight and knowledge into the matter of project management regarding this thesis helped very much. Additionally he supported me in the identification of the thesis topic and the creation of the papers structure. Furthermore, I would like to thank my supervisors Barbara Hansen

and Anton Obermaier at the company side who helped me getting into the most important topics of this thesis and provided me the best possible advice.

REFERENCES

[1] M. Bornschlegl, "Eine Methode zur Energieplanung für Fügeverfahren im Karosseriebau," 4 2016.
[2] H. Fahrenwaldt, V. Schuler, and J. Twrdek, *Praxiswissen Schweißtechnik*, 4th ed. Wiesbaden, Germany: Springer Fachmedien, 2011.
[3] H. Kerl, "Ein Beitrag zur Standmengenerhöhung der Elektrodenkappen beim Widerstandspunktschweißen," 2017.
[4] H.-H. Braess and U. Seiffert, *Handbuch Kraftfahrzeugtechnik*, 8th ed. Wiesbaden, Germany: Springer Fachmedien, 2021.
[5] J. Fridley and J. Jorgensen, "Benchmarking: a process basis for teaching design," *Frontiers in Education Conference*, vol. 2, pp. 960–967, 12 1997.
[6] W. Karl and K. Roman, *PQM - Prozessorientiertes Qualitätsmanagement*, 3rd ed. München, Germany: Carl Hanser Verlag, 2006.
[7] W. Lankford, "Benchmarking: Understanding the basics," *The Coastal Business Journal*, vol. 1, pp. 57–62, 1 2000.
[8] R. Kaiser, *Qualitative Experteninterviews*, 3rd ed. Wiesbaden, Germany: Springer Fachmedien, 2014.

Bernhard Spornraft is currently enrolled in the Masters program "Mechatronics and Smart Technologies" at the Management Center Innsbruck. He has a bachelors degree in mechanical engineering and successfully finished an apprenticeship as mechatronic technician.

Entwicklung und Umsetzung eines flexiblen und modularen Automatisierungskonzepts zur robotergestützten Maschinenbestückung

Marcel Stifter, Armin Lechner (Betreuer) und Benjamin Massow (Betreuer)

***Kurzfassung*— Der strikt hierarchische Programmaufbau vieler Automatisierungsanlagen führt bereits bei geringen Prozessänderungen zu enormen Planungs- und Programmieraufwänden. Diese Aufwände haben lange Stillstandszeiten und erhöhte Kosten zur Folge. Besonders klein- bis mittelständische Unternehmen sind von diesem Problem betroffen, da ihre Produktion vorrangig auf geringe Stückzahlen ausgelegt ist und dadurch vermehrt Prozessänderungen anfallen. Ihrer Forderung nach Flexibilität stehen nahezu keine geeigneten Automatisierungslösungen zur Verfügung. Anhand einer flexiblen Programmierung kann eine anpassbare Automatisierungsanlage entwickelt werden, welche anschließend den Unternehmen als neuartige Gesamtlösung zur Verfügung steht. Überprüft wird diese Annahme durch die Realisierung einer mobilen modularen Maschinenbestückungsanlage. Der Aufbau des Programms erfolgt objektorientiert. Jedes Modul des Systems erhält eine vorgegebene Struktur und kann mit beliebiger, dazu passender Hardware realisiert werden. Dank dieses Systemaufbaus ist es möglich, ohne große Anpassungen am Steuerungsprogramm, verschiedenste Hardwarekomponenten von unterschiedlichen Herstellern in das System einzubinden und anzusteuern. Die Ergebnisse dieser Arbeit zeigen, dass anhand einer flexiblen Bestückungsanlage mehrere Werkzeugmaschinen über verschiedene Kommunikationsprotokolle schnell und kostengünstig automatisiert werden können, ohne dabei große Programmieraufwände bei Prozessänderungen oder Maschinenwechsel zu generieren. Dank einer adaptiven Programmstruktur können moderne Anlagen flexibel gestaltet und neuartige Lösungskonzepte für dynamische Automatisierungssysteme entwickelt werden.**

***Schlagwörter*— Flexible Automatisierung, Objektorientierte Programmierung, Roboterplattform, Modulare Maschinenbestückung, Smart Factory.**

I. Einleitung

VIELE moderne Automatisierungsanlagen basieren nach wie vor auf alten Steuerungskonzepten, wobei starr programmierte Ablaufketten kontinuierlich abgearbeitet werden. Dieses klassische Automatisierungskonzept stößt bei häufigen System- und Prozessänderungen jedoch schnell an seine Grenzen. Bereits kleine Anpassungen können enorme Planungs- und Programmieraufwände hervorrufen, die wiederum in langen Stillstandszeiten, erhöhtem Personalbedarf und hohen Kosten resultieren. Klein- bis mittelständische Unternehmen (KMU's) sind maßgeblich von diesem Problem betroffen, da sie vorrangig individualisierte Produkte mit kleinen Losgrößen produzieren und dadurch Prozessänderungen vermehrt auftreten. Herkömmliche Automatisierungslösungen bieten meist keine ausreichende Flexibilität, wodurch selbst einfache Produktionsschritte, wie beispielsweise eine Maschinenbestückung, manuell durchgeführt werden müssen [1]. Aktuelle Marktforschungsstudien zeigen einen klaren Trend in Richtung individualisierter Produkte [2], weshalb viele Unternehmen eine geeignete Automatisierungslösung für eine flexible Produktion fordern. Dennoch ist nahezu kein ökonomisch sinnvolles Konzept für eine flexible Automatisierung von Werkzeugmaschinen am Markt erhältlich.

Ziel dieser Arbeit ist es, ein Gesamtkonzept für eine flexible und modulare Automatisierung von Werkzeugmaschinen für kleine Losgrößen zu entwickeln und zu realisieren. Die Zielgruppe des Konzepts sollen dabei KMU's darstellen. Das ausgearbeitete Gesamtkonzept wird in Form eines Funktionsprototyps umgesetzt und getestet. Dabei wird auf die Realisierung des elektrischen Aufbaus sowie die Programmerstellung genauer eingegangen. Für eine aussagekräftige Konzeptevaluierung wird der Prototyp anhand einer Use Case Analyse auf dessen Funktionalität überprüft und die Vorteile einer flexiblen Automatisierung belegt.

M. Stifter studiert am Studiengang Mechatronik, MCI, Innsbruck, Österreich, und arbeitet beruflich an der Entwicklung von Elektromotoren im Automobilsektor, GKN Driveline, Bruneck, Italien, e-mail: sm4848@mci4me.at.

A. Lechner arbeitet am Studiengang Mechatronik, MCI, Innsbruck, Österreich, e-mail: armin.lechner@mci.edu.

B. Massow arbeitet am Studiengang Mechatronik, MCI, Innsbruck, Österreich, e-mail: benjamin.massow@mci.edu.

II. Gesamtkonzept

A. Entwicklung des Gesamtkonzepts

Bevor ein geeignetes Gesamtkonzept für eine neuartige Bestückungsanlage entwickelt werden kann, muss eine Problemanalyse durchgeführt werden. Anhand dieser Analyse werden vier große Problemfelder von KMU's ermittelt: Technik, Finanzen, Zeit und Anwendbarkeit. Die neue Gesamtlösung muss eine umfassende Produkt- und Herstellerunabhängigkeit aufweisen, da verschiedenste Bauteile an unterschiedlichsten Werkzeugmaschinen hergestellt werden. Die entstehenden Kosten für die Realisierung des Systems sind während der Entwicklungsphase stets im Auge zu behalten. Fallen zu hohe Kosten an, so wird das System für KMU's unbrauchbar. Weiters muss eine möglichst kurze Umrüstzeit bei Produkt- und Maschinenwechsel erzielt werden, sodass die Bestückungsanlage selbst bei geringen Stückzahlen verwendet werden kann. Ein wichtiger Aspekt ist dabei die einfache Bedienung der Anlage. Fachpersonal ist in jedem Unternehmen nur begrenzt vorhanden. Aufgrund dieser Tatsache muss die Anlagenbedienung so intuitiv wie möglich gestaltet werden, damit jegliches Personal das System starten und betreuen kann.

Um die zuvor genannten Probleme und Anforderungen in einem flexiblen Automatisierungssystem abdecken zu können, muss das Prinzip der Modularität angewandt werden. Die neuartige Gesamtlösung soll nicht nur eine einfache Bestückungsanlage realisieren, sondern KMU's eine fortgeschrittene Automatisierungslösung für die gesamte Produktion bieten. Das System besteht dabei aus verschiedensten Modulen, die zu jedem Zeitpunkt hinzugefügt, ersetzt und entfernt werden können. Jedes Modul umfasst eine gewisse Anzahl an unterschiedlicher Hardware, aus der individuell eine passende Komponente gewählt werden kann. Diese Eigenschaft ermöglicht es, sich Kundenanforderungen sowohl in technischen als auch finanziellen Belangen bestmöglich anzupassen. Abbildung 1 zeigt alle relevanten Module, die in Bezug auf eine flexibel automatisierte Bestückungsanlage ermittelt werden.

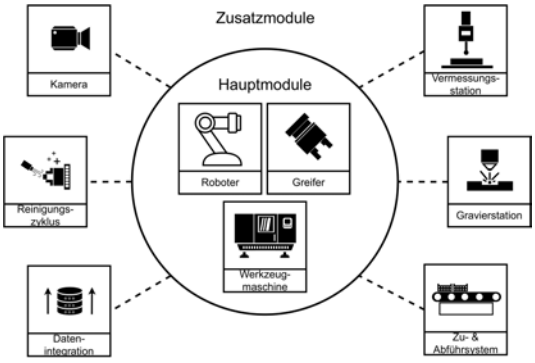


Abbildung 1. Ermittelte Module für das Gesamtkonzept einer flexiblen Bestückungsanlage von Werkzeugmaschinen.

Die Werkzeugmaschine, der Roboter und der Greifer bilden die Hauptmodule des Systems. Sie werden zwangsläufig für den Bestückungsprozess benötigt und sind daher für das System unerlässlich. Anhand der Modularisierung von Werkzeugmaschinen kann eine allgemeine und standardisierte Ansteuerung realisiert werden, wodurch eine herstellerunabhängige Maschineneinbindung in das System ermöglicht wird. Das Robotermodul wird in Form eines kollaborativen Roboters umgesetzt, sodass die Bestückungsanlage mobil ausgeführt werden kann. Weiters wird dadurch eine einfache Bedienung über manual guiding und ein benutzerfreundliches Human Machine Interface (HMI) erzielt. Der Greifer des Systems muss über eine eigene Intelligenz und Logikeinheit verfügen, sodass fortgeschrittene Prozessüberwachungen wie Spanndetektionen und error handlings realisiert werden können. Das Kameramodul kann sowohl für Qualitätskontrollen als auch für Späneerkennungen verwendet werden. Der Reinigungszyklus steht für die Bauteil- und Spannmittelreinigung zur Verfügung. Mittels einer Datenintegration, Vermess- und Gravierstation kann eine nahtlose Überprüfung aller produzierten Bauteile realisiert werden. Anhand eines automatisierten Zu- und Abführsystems von Bauteillagern wird eine signifikante Produktionsvolumenerhöhung bei kurzen Durchlaufzeiten ermöglicht.

B. Aufbau des Gesamtsystems

Zunächst muss ein geeigneter Hardwareaufbau für das Gesamtsystem definiert werden. Hierfür wird auf ein bereits bestehendes Konzept [3] zurückgegriffen und dieses im Detail ausgearbeitet. Der entwickelte Systemaufbau ist in Abbildung 2 grafisch aufgezeigt.

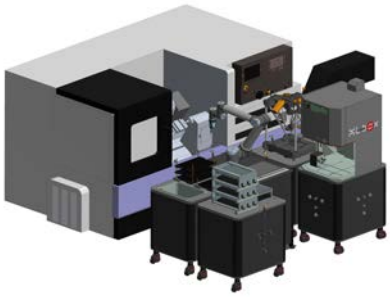


Abbildung 2. Hardwareaufbau des Gesamtkonzepts für eine flexible Werkzeugmaschinenbestückung. Adaptiert nach [3].

Die Robotereinheit bildet das Hauptmodul des Gesamtsystems. Sie wird über eine Steckverbindung mit der Werkzeugmaschine verbunden und dadurch mit Strom, Druckluft und Daten versorgt. Anschließend bildet das Hauptmodul die zentrale Versorgungseinheit für alle weiteren Module. Jedes Zusatzmodul wird über einen Mechanismus und eine Steckverbindung mit dem Hauptmodul gekoppelt und folglich in das System eingebunden. Um eine möglichst große Flexibilität erzielen zu können, muss jedes Modul mobil ausgeführt werden. Dies kann über ein standardisiertes mechanisches Grundgerüst realisiert werden.

Für die Entwicklung einer geeigneten Ansteuerungsstruktur wird erneut ein bereits bestehendes Konzept [4] herangezogen und dieses weiter ausgearbeitet. Die entwickelte Ansteuerungsstruktur ist in Abbildung 3 ersichtlich.

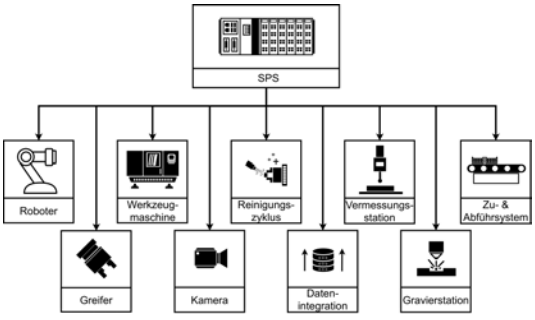


Abbildung 3. Ansteuerungsstruktur des Gesamtsystems für eine flexible Werkzeugmaschinenbestückung. Adaptiert nach [4].

Die Ansteuerung der Module erfolgt anhand einer übergeordneten speicherprogrammierbaren Steuerung (SPS). In diesem System fungiert die SPS als Moderator und alle

weiteren Module als Teilnehmer [4]. Aufgrund der Tatsache, dass verschiedene Hardwarekomponenten mit unterschiedlichen Schnittstellensignalen in das System eingebunden werden, muss jeder Teilnehmer gekapselt realisiert werden. Der Moderator triggert Funktionen von Teilnehmern und überwacht anschließend deren Ablauf [4]. Alle Teilnehmerfunktionen arbeiten autonom und sind direkt auf der Logikeinheit der Teilnehmer implementiert. Der Moderator erhält als Rückmeldung von jedem Teilnehmer den Status der Operation sowie eine Bestätigung, sobald sie vollendet ist. Dank dieses Aufbaus kann der Steuerungscode der SPS in einer allgemeinen Form realisiert werden, wodurch bei Modulwechsel nahezu kein zusätzlicher Programmieraufwand entsteht. In der vorhandenen Konstellation übernimmt die SPS die Funktion der Ablaufsteuerung sowie der Überwachung des gesamten Produktionsprozesses [4].

Anhand der entwickelten Ansteuerungsstruktur treten im System zwei unterschiedliche Kommunikationsarten auf, direkt und indirekt. Eine direkte Kommunikation ist eine reale Schnittstellenkommunikation, die ausschließlich zwischen Moderator und Teilnehmern aufgebaut wird. Die SPS kann mittels Hardwareerweiterungen nahezu jede Kommunikationsschnittstelle abdecken, wodurch eine kommunikationsunabhängige Teilnehmeransteuerung realisiert werden kann. Verschiedene Teilnehmer können dabei gleichzeitig über unterschiedliche Schnittstellen mit dem Moderator kommunizieren und Informationen austauschen. Eine indirekte Kommunikation wird zwischen zwei Teilnehmern aufgebaut und erfolgt über die SPS. Bei komplexen Funktionen müssen Teilnehmer untereinander kommunizieren, um Operationen korrekt ausführen zu können. Jeder Teilnehmer kann eine Anforderung an die SPS senden, sodass benötigte Teilnehmerfunktionen getriggert und Informationen ausgetauscht werden. Dank dieser indirekten Kommunikation können Teilnehmer untereinander kommunizieren, ohne über eine einheitliche Kommunikationsschnittstelle zu verfügen.

Im Anschluss an die Definition des Systemaufbaus muss ein Ablauf für die Bedienung der Anlage identifiziert werden. Der Bediener muss in der Lage sein die Anlage zu kontrollieren, einzurichten, zu parametrisieren und zu starten. In Bezug auf das Gesamtsystem wird folgender Ablauf für die Bedienung der Anlage bestimmt:

- Sichere Initialisierung der Analge beim Einschaltvorgang
- Wahl der vorhandenen Teilnehmer des Systems
- Einrichten der Anlage im Handbetrieb
- Parametrisierung einer Ablaufkette für flexible Automatisierungsprozesse
- Rezepturverwaltung zur Wiederverwendung bereits implementierter Abläufe

Die Anlage muss in einem sicheren Not-Aus Zustand initialisieren, da keine Information über den aktuellen Status der Teilnehmer vorhanden ist. Im Anschluss an den Einschaltvorgang erfolgt die Teilnehmerauswahl, wobei eine aktive Kommunikation zu allen Modulen aufgebaut wird. Mittels eines Handbetriebs muss der Bediener in der Lage sein die Anlage einzurichten, zu kalibrieren und in einen definierten Zustand zu versetzen. Aufgrund der Tatsache, dass der Produktionsprozess frei gestaltet werden kann, muss die Anlage über eine parametrisierbare Ablaufkette [4] verfügen, sodass der Produktionsablauf an die gegebenen Anforderungen angepasst werden kann. Im Zuge dieser freien Parametrisierbarkeit muss eine Rezepturverwaltung eingeführt werden, sodass bereits funktionierende Abläufe gespeichert und erneut gestartet werden können [4].

III. Umsetzung des Gesamtsystems

A. Realisierung eines Funktionsprototyps

Das entwickelte Gesamtsystem wird in Form eines Funktionsprototyps realisiert, sodass dessen Funktionalität überprüft werden kann. Dabei wird jedoch nicht direkt das gesamte System aufgebaut, sondern vielmehr die wichtigsten und nützlichsten Module umgesetzt. Abbildung 4 zeigt alle Module, die im realisierten Funktionsprototyp enthalten sind.

Abbildung 4. Übersicht der Module, die als Prototyp realisiert werden.

Die Entwicklung des Funktionsprototyps wird in vier große Themenbereiche unterteilt: Die Realisierung des mechanischen Aufbaus [5], die Auslegung eines Sicherheitskonzepts [6], [7], die Entwicklung einer Späneerkennung sowie eines Reinigungszyklus [8] und die Entwicklung des elektrischen Aufbaus sowie der dazugehörigen Programmierung. Der letztgenannte Themenbereich wird in weiterer Folge im Detail ausgearbeitet und beschrieben.

B. Elektrischer Aufbau

Zunächst müssen geeignete Hardwarekomponenten für alle vorhandenen Module des Funktionsprototyps definiert werden. In Tabelle I sind die verwendeten Komponenten des Systems aufgelistet.

Tabelle I
Übersicht über alle verwendeten Hardwarekomponenten für die Umsetzung des Funktionsprototyps.

Modul	Hersteller	Artikel
Werkzeugmaschine	*Fanuc*	*Robodrill α-D21MiB5*
Roboter	*Fanuc*	*CRX 10iA/L*
Greifer	*Schunk*	*EGL 90-PN*
Kamera	*Cognex*	*D905*
Reinigungsbox	-	-

Die erste Werkzeugmaschine des Systems bildet die *Fanuc Robodrill α-D21MiB5*, da eine solche Maschine für eine längere Testphase zur Verfügung steht. Die Wahl des Roboters fällt auf den kollaborativen Roboter *Fanuc CRX-10iA/L*, da er über eine große Reichweite verfügt und Lasten bis zu 10 kg tragen kann. Als Werkzeug des Roboters wird der Greifer *Schunk EGL 90-PN* definiert. Dieser besitzt eine maximale Spannkraft von 600 N und einen maximalen Hub

von 42.5 mm. Die Kamera und Reinigungsbox werden von [8] übernommen und in das System eingebunden.

Im Anschluss an die Hardwareauswahl muss ein passender Schaltschrank für die Versorgung und Ansteuerung der Komponenten entwickelt werden. Sowohl der Schaltschrank als auch der Robotercontroller werden im Innenraum des zentralen Hauptmoduls verbaut. Bei der Realisierung des Schaltschranks werden die Normen *DIN EN 60204* [9], *DIN EN 61439-1* [10], *DIN EN 61439-2* [11], *DIN EN 61355-1* [12] sowie *DIN EN 81346-2* [13] berücksichtigt und größtenteils bereits in der Prototypenphase eingehalten. Die Versorgung erfolgt, wie vom Gesamtkonzept gefordert, über eine *Harting* Steckverbindung. In ihr ist eine 230 V Spannungsversorgung, ein Druckluftanschluss sowie eine Profibus und Profinet Kommunikationsschnittstelle enthalten. Der Robotercontroller *Fanuc R-30iB Mini* wird über einen 10 A Automat abgesichert. Aufgrund der Tatsache, dass mehrere Teilnehmer eine 24 V Versorgung benötigen, wird die Ausgangsspannung eines 10 A Netzgerät über mehrere elektronische Sicherungen abgesichert und auf drei Kreise aufgeteilt. Der erste Kreis versorgt dabei alle internen Komponenten, der zweite alle externen und der dritte alle Leistungsabgänge. Die Profinet Datenverteilung erfolgt anhand eines einfachen achtfach-Ethernet Switchs. Über einen verbauten WLAN-Router kann eine Kommunikation zu einem externen Netzwerk realisiert werden. Die Steuerung des gesamten Systems erfolgt anhand einer SPS *Beckhoff CX5130*. Dabei handelt es sich um einen Industrial Personal Computer (IPC) mit dem Betriebssystem *Windows 7 Embedded*. Die SPS verfügt über eine achtfach digitale Ausgangsklemme, eine Profibus Masterklemme sowie zwei Safety Klemmen mit jeweils vier Eingängen und einem Ausgang. Diese Ausgänge werden mittels Safety Relais erweitert, sodass potentialfreie Kontakte zur Verfügung stehen. Der Sicherheitskreis der Anlage wird anhand eines Safety Projekts in der SPS realisiert.

Der vollendete Aufbau des Funktionsprototyps ist in Abbildung 5 ersichtlich. Das mechanische Grundgerüst wird aus [5] übernommen. Neben dem Hauptmodul befindet sich ein einfaches provisorisches Lagermodul aus Holz.

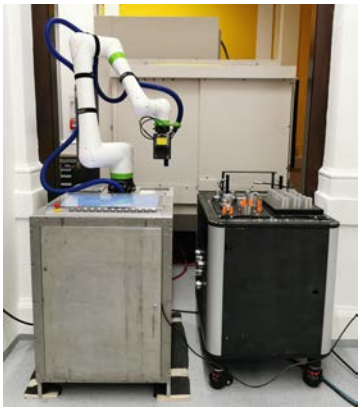


Abbildung 5. Aufbau des vollendeten Funktionsprototyps mit Lagermodul.

C. Entwicklung des Steuerungsprogramms

Die Implementierung des Steuerungsprogramms erfolgt in der Software *TwinCat XAE Shell* von *Beckhoff*. Alle verwendeten Softwaremodule und Versionen sind in Tabelle II aufgelistet. Das gesamte Programm wird in Form von Structured Text (ST) realisiert.

Tabelle II
ÜBERSICHT ÜBER DIE VERWENDETEN SOFTWAREVERSIONEN FÜR DIE ERSTELLUNG DES STEUERUNGSPROGRAMMS.

Softwarepaket	Version
TwinCat 3 Engineering	3.1.4024.17
TwinCat 3 Runtime	3.1.4024.17
TE2000 HMI Engineering	1.12.746.3
TF2000 HMI Server	1.12.746.3

Grundlage für die Erstellung des neuen Steuerungsprogramms bildet ein bereits entwickelter Konzeptcode [4]. Die Struktur dieses Codes wird übernommen, das gesamte Programm jedoch gänzlich erneuert und ausgearbeitet.

Als erster Schritt der Programmerstellung muss ein möglicher Ablauf der Maschinenbestückung definiert und dieser in einzelne Operationen unterteilt werden. Die ermittelten Operationen werden über verschiedene Teilnehmerfunktionen realisiert und bilden die Grundlage für die Entwicklung der Bausteine, die in weiterer Folge dem Bediener für die Parametrisierung der Ablaufkette zur Verfügung stehen. Abbildung 6 zeigt die Unterteilung des Bestückungsprozesses in die definierten Operationen.

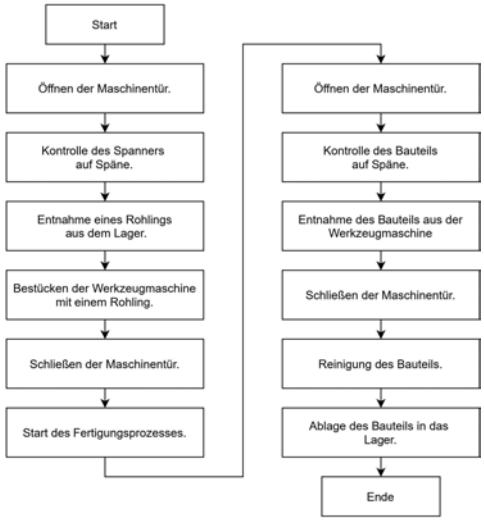


Abbildung 6. Unterteilung des Bestückungsprozesses in die Operationen.

Zunächst wird im Hauptprogramm eine Programmstruktur für die Steuerung der Anlage ausgearbeitet. Anschließend werden alle Teilnehmer in das System eingebunden und ein standardisierter Aufruf der Teilnehmerfunktionen entwickelt.

D. Aufbau des Hauptprogramms

Das Hauptprogramm kann allgemein in drei Bereiche unterteilt werden: Die Teilnehmerauswahl, den Handbetrieb und den Automatikbetrieb.

Die Teilnehmerauswahl erfolgt anhand einer einfachen Switch/Case Anweisung [4]. Jedem Modul wird eine Enumeration-Variable zugewiesen, welche den entsprechenden Teilnehmer repräsentiert [4]. Besitzt diese Variable den Wert 0, so ist das entsprechende Modul im System nicht vorhanden. Nimmt die Variable den Wert eines Teilnehmers an, so wird der teilnehmerspezifische Funktionsbaustein aufgerufen und ein dazugehöriges Interface beschrieben [4]. Wie bei jeder objektorientierten Programmierung bildet das Interface eine Klasse, die über eine bestimmte Anzahl an Methoden verfügt. Diese Methoden entsprechen dabei den Teilnehmerfunktionen der Hardware. Als Teil einer Klasse muss jeder Teilnehmer alle Methoden des Interfaces abdecken. Das Interface selbst fungiert als Pointer, welcher auf den Speicherplatz des übergebenen Funktionsbausteins zeigt. Der Aufruf der Teilnehmerfunktionen kann dadurch über das Interface erfolgen, wodurch der Steuerungscode in einer allgemeinen Form realisiert werden kann. Die genaue Methodenimplementierung erfolgt in den teilnehmerspezifischen Funktionsbausteinen. Der Ablauf der Methoden muss teilnehmerübergreifend nicht identisch sein. Anhand dieser offenen Implementierungsvariante können hardwaregebundene Besonderheiten, wie beispielsweise Signaltimecharts, abgedeckt werden, ohne die Funktionalität anderer Teilnehmer zu beeinflussen.

Der Handbetrieb bildet den Einrichtbetrieb der Anlage und wird für Test- und Inbetriebnahmezwecke verwendet. Nach jedem Neustart muss die Anlage auf ihre Funktionalität überprüft und in eine sichere Grundstellung gebracht werden. Der Handbetrieb ermöglicht dabei das direkte Starten von Teilnehmerfunktionen, ohne den gesamten Produktionsablauf parametrisieren zu müssen. Die Ansteuerung der Methoden erfolgt anhand einer einfachen Startvariable. Während eine Methode abgearbeitet wird, kann keine weitere gestartet werden. Nach Vollendung des Methodenablaufs wird die Startvariable automatisch resettiert und die Anlage für die weitere Bedienung freigegeben. Dadurch können alle Funktionen der Teilnehmer vor Produktionsbeginn überprüft, kalibriert und gegebenenfalls nachjustiert werden.

Der Automatikbetrieb realisiert den normalen Produktionsbetrieb der Anlage. Jeder Produktionsablauf beginnt mit einer Startfreigabe des Bedieners. Anschließend wird eine einmalige Initialisierungsroutine durchlaufen, welche die Anlage über eine fix definierte Methodenabfolge in einen sicheren Startzustand befördert. Sobald dieser Zustand erreicht ist, beginnt die Anlage mit der Abarbeitung aller Methoden laut einer benutzerdefinierten State Machine. Bei der State Machine handelt es sich um ein Enumeration-Array, worin jede Teilnehmermethode als spezifischer Zahlenwert definiert ist [4]. Der Bediener füllt dieses Array mit Werten und bestimmt dadurch den genauen Produktionsablauf. Der Zahlenwert 0 besitzt dabei eine Sonderstellung, da er das Ende der parametrisierten Ablaufkette symbolisiert [4]. Bei jeder Vollendung einer Teilnehmermethode wird ein Signal für die Zustandsänderung der State Machine gesetzt, wodurch der nächste Arrayeintrag aufgerufen wird [4]. Dieser Vorgang wiederholt sich bis das Ende der parametrisierten State Machine erreicht ist. Sollten weitere Bauteile benötigt werden, so beginnt der Durchlauf erneut und ein Bauteilcounter wird erhöht. Nachdem der Bauteilcounter die Anzahl der gewünschten Bauteile erreicht, wird eine Endroutine abgearbeitet. Dabei wird wie zuvor in der Initialisierungsroutine eine fixe Methodenabfolge durchlaufen, worüber die Anlage in einen sicheren Endzustand befördert wird. Nach erfolgreichem Abschluss der Endroutine kann der Bediener das Produktionsende bestätigen und die Anlage für die nächste Produktion freigeben.

E. Teilnehmereinbindung und Methodenimplementierung

Im Anschluss an die Erstellung des Hauptprogramms muss ein hardwareseitiger Kommunikationsaufbau zu allen Komponenten des Systems realisiert werden. Jeder Teilnehmer wird als Profinet Device angelegt und parametrisiert. Dabei muss ein eindeutiger Stationsname, eine IP-Adresse sowie ein passendes GSDML-File gewählt werden. Im Zuge dieses Kommunikationsaufbaus müssen alle Schnittstellensignale klar definiert und in den teilnehmerspezifischen Funktionsbausteinen implementiert werden. Der Greifer und die Kamera verfügen bereits über eine standardisierte Schnittstellenkommunikation, weshalb diese direkt übernommen werden kann. Beim Roboter und der Werkzeugmaschine ist die Kommunikation individuell konfigurierbar, wodurch eine manuelle Schnittstellenstandardisierung erforderlich ist. Diese Schnittstellen müssen anschließend bei allen Teilnehmern derselben Modulart identisch umgesetzt werden, sodass der zukünftige Programmieraufwand verringert wird.

Die Methodenimplementierung ermöglicht die Ansteuerung der Teilnehmer über die SPS. Mittels Methoden können Teilnehmerfunktionen gestartet und Informationen ausgetauscht werden. Bei der Methodenimplementierung wird zwischen zwei Arten von Methoden unterschieden, den allgemeinen und den modulspezifischen Methoden.

Allgemeine Methoden sind standardisiert und werden für definierte Aufgaben verwendet. Darin sind Informations- und Ansteuerungsmethoden enthalten, die für die korrekte Funktion des Systems unerlässlich sind. Sie beinhalten Warnsignale, Kommunikationskontrollen, Statusmeldungen und Abbruchbefehle. Die Implementierung der Informationsmethoden muss bei jedem Teilnehmer eines Moduls zwingend identisch realisiert werden.

Modulspezifische Methoden sind nur bei Teilnehmern derselben Modulart vorhanden. Diese Methoden werden für die Ansteuerung der Teilnehmerfunktionen verwendet, wodurch Hardwareoperationen ausgeführt werden. Der Ansteuerungsablauf ist standardisiert und folgt einem definierten Schema. Dieser Ablauf ist in Abbildung 7 in allgemeiner Form dargestellt. Über das Hauptprogramm wird der Methodenablauf zyklisch wiederholt bis der Teilnehmer das Funktionsende bestätigt oder ein Abbruch erfolgt.

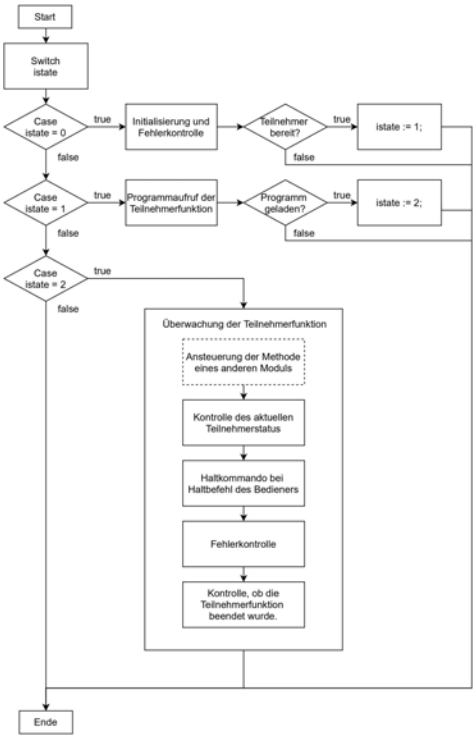


Abbildung 7. Allgemeines Schema des Methodenablaufs für die Ansteuerung einer Teilnehmerfunktion. Der strichlierte Block symbolisiert die Ansteuerung von Teinehmermethoden eines anderen Moduls, wodurch eine Zusammenarbeit zwischen Teilnehmern realisiert werden kann.

IV. Evaluierung des Gesamtsystems

Für eine Evaluierung des Gesamtsystems muss der entwickelte Funktionsprototyp anhand eines geeigneten Use Cases auf seine Funktionalität überprüft werden. Dabei soll ein möglichst realistisches Szenario aus Sicht eines KMU betrachtet werden. Entscheidet sich ein Unternehmen das System zu erwerben, so werden alle benötigten Module mit den gewünschten Hardwarekomponenten realisiert und vor Ort in Betrieb genommen. Im Anschluss werden vorrangig keine neuen Hardwaremodule ersetzt, sondern vielmehr möglichst viele vorhandene Werkzeugmaschinen in das System eingebunden. Anhand dieser Tatsache ist es naheliegend, die Modularität des Systems an verschiedenen Werkzeugmaschinen zu testen, da dies in der Praxis höchste Relevanz besitzt.

Der entwickelte Funktionsprototyp wird als Use Case an vier möglichst unterschiedlichen Werkzeugmaschinen getestet, sodass dessen Flexibilität, Modularität und Funktionalität belegt werden kann. Die vier vorhandenen Werkzeugmaschinen sind in Tabelle III aufgelistet. Zunächst wird der Erstinbetriebnahmevorgang aller Maschinen verglichen und genau analysiert. Anschließend wird der Einricht- und Umrüstprozess für eine Produktionsautomatisierung an zwei Werkzeugmaschinen untersucht und gegenübergestellt. Den Abschluss der Analyse bildet eine reale Testproduktion an einer Werkzeugmaschine, worüber die Prozesssicherheit während eines Dauerbetriebs ermittelt wird.

Tabelle III
Übersicht aller Werkzeugmaschinen, die in das System eingebunden werden.

Hersteller	Typ	Steuerungstyp	Schnittstelle
Fanuc	*Robodrill α-D21MiB5*	*Serie 31i- Model B*	Profinet
Tsugami	*BH38E*	*Serie 31i- Model A*	Profibus
Tsugami	*MO8SYE*	*Serie 0i-TD*	Profibus
Doosan	*Puma 2600SY II*	*Serie i*	Profinet

A. Evaluierung der Erstinbetriebnahmen

Bei jeder Erstinbetriebnahme einer Werkzeugmaschine müssen fünf Arbeitsschritte durchgeführt werden. Zunächst wird der Teilnehmer in das System eingebunden und alle Kommunikationssignale nach standardisierten Schnittstellenvorgaben konfiguriert. Anschließend müssen alle Methoden für die Ansteuerung der Teilnehmerfunktionen implementiert und in Folge sämtliche Signal- und Ansteuerungschecks durchgeführt werden. Den Abschluss der Erstinbetriebnahme bildet eine Kontrolle der Erreichbarkeit aller Roboterpositionen, die für den Bestückungsprozess benötigt werden. Der entstandene zeitliche Aufwand für die Erstinbetriebnahme von jeder Werkzeugmaschine wird in Abbildung 8 grafisch aufgezeigt und gegenübergestellt.

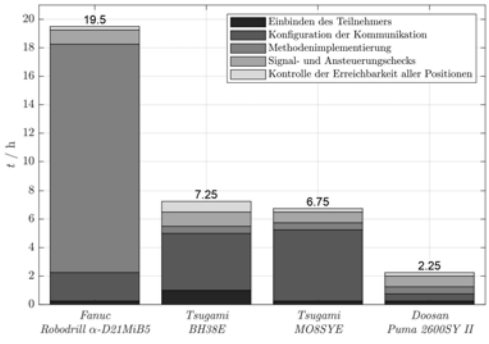


Abbildung 8. Gegenüberstellung des zeitlichen Aufwands für die Erstinbetriebnahme von vier unterschiedlichen Werkzeugmaschinen.

Der zeitliche Aufwand für die Erstinbetriebnahme der Werkzeugmaschine *Fanuc Robodrill α-D21MiB5* ist wesentlich höher als bei allen weiteren Erstinbetriebnahmen. Dies ist auf den hohen Programmieraufwand der Methodenimplementierung zurückzuführen. Sie wird als erste Werkzeugmaschine in das System aufgenommen, wodurch der gesamte Ansteuerungscode an ihr entwickelt werden muss. Dank der standardisierten Ansteuerungsstruktur des Systems kann diese Methodenimplementierung anschließend bei allen weiteren Werkzeugmaschinen übernommen werden, wodurch eine hohe Zeitersparnis erzielt wird. Bei beiden Werkzeugmaschinen des Herstellers *Tsugami* ist ein erhöhter Zeitaufwand für die Konfiguration der Kommunikationsschnittstelle ersichtlich.

Dieser Hersteller verfügt in seinen Produkten über keine standardisierte Programmstruktur, wodurch jegliches Ansteuerungssignal einzeln im Steuerungscode der Werkzeugmaschine gesucht werden muss. Dadurch entsteht bei jeder Erstinbetriebnahme ein erhöhter Aufwand, dessen Auswirkung nur schwer im Vorfeld abgeschätzt werden kann. Die Erstinbetriebnahme der Werkzeugmaschine *Doosan Puma 2600SY II* zeigt eindeutig, welche Aufwandsreduktion anhand der Modularisierung von Werkzeugmaschinen ermöglicht wird. Grundsätzlich gilt, je mehr Maschinen in das System eingebunden werden, desto größer und signifikanter wird die Zeitersparnis durch Codewiederverwertung und Schnittstellenstandardisierung.

B. Evaluierung des Einricht- und Umrüstprozesses

Die Dauer des Einricht- und Umrüstprozesses ist eine wichtige Kenngröße, die beschreibt wie viel Zeit für das Anpassen des Automatisierungsprozesses benötigt wird. Anhand dieser Kenngröße und der geforderten Bauteilanzahl kann der Bediener entscheiden, ob sich der entstehende Aufwand mit dem erhaltenen Nutzen der Produktionsautomatisierung deckt. Beide Prozesse werden an zwei möglichst unterschiedlichen Werkzeugmaschinen umgesetzt und die entstehenden Aufwände verglichen. Die Analyse erfolgt an den Werkzeugmaschinen *Fanuc Robodrill α-D21MiB5* und *Doosan Puma 2600SY II*.

Der Einrichtprozess tritt bei jeder Erstinbetriebnahme und bei jedem Maschinenwechsel auf. Dabei müssen sowohl der Automatisierungsablauf als auch alle benötigten Operationen neu eingelernt werden. Aufgrund der Tatsache, dass die Werkzeugmaschine *Doosan Puma 2600SY II* über einen eigenen Stangenlader verfügt, reduziert sich der Bestückungsvorgang auf eine Bauteilentnahme. Der benötigte zeitliche Aufwand für das Einlernen aller Operationen ist in Tabelle IV aufgelistet.

Tabelle IV
ZEITAUFWAND FÜR DAS EINRICHTEN ALLER OPERATIONEN AN DER *Fanuc Robodrill α-D21MiB5* UND DER *Doosan Puma 2600SY II*.

Operation	*Robodrill α-D21MiB5* t / h	*Puma 2600SY II* t / h
Rohteil aufnehmen	1.00	-
Rohteil bestücken	0.50	-
Fertigteil entnehmen	0.25	0.50
Bauteil reinigen	0.25	0.25
Fertigteil ablegen	0.50	1.00
Späneerkennung Bauteil	4.00	4.00
Späneerkennung Spanner	4.00	4.00
Automatische Tür öffnen	0.00	0.00
Automatische Tür schließen	0.00	0.00
CNC Start	0.00	0.00
Summe	10.50	9.75

Beim Einrichtprozess müssen grundsätzlich zwei Arbeitsschritte durchgeführt werden: Das Einlernen von Roboterpositionen und das Entwickeln von Kamerajobs für die Späneerkennung. Jede Roboterposition muss neu definiert und an die vorhandene Werkzeugmaschine angepasst werden. Dieser Aufwand wird durch eine geschickte Wahl von Roboterpositionen verringert, indem standardisierte Positionen definiert und diese Programmübergreifend verwendet werden. Den größten Aufwand des Einrichtprozesses bildet das Entwickeln von Kamerajobs. Bei jedem implementierten Kamerajob müssen ungefähr 100 unterschiedliche Bilder mit Spänen an der jeweiligen Komponente aufgezeichnet werden. Der zeitliche Aufwand für die Erstellung eines Kamerajobs beträgt $1.5\,\mathrm{h}$. Anhand der Tatsache, dass bei jeder Späneerkennungsposition mindestens zwei Kamerajobs benötigt werden, erhöht sich dieser Aufwand auf $3\,\mathrm{h}$. Zusätzlich zu den erstellten Kamerajobs müssen alle Roboterpositionen für den Reinigungszyklus definiert werden, wodurch sich der angegebene Zeitaufwand für die gesamte Operation ergibt. Bezogen auf die erhobenen Daten beläuft sich der maximale Zeitaufwand für den gesamten Einrichtprozess auf ungefähr 1 bis 2 Arbeitstage.

Der Umrüstprozess muss bei jedem Bauteilwechsel durchgeführt werden und entspricht einem verkürzten Einrichtprozess. Beim Umrüstprozess ist der Automatisierungsablauf bekannt und die meisten Operationen sind bereits größtenteils eingelernt. Tabelle V zeigt den entstehenden zeitlichen Aufwand für das Nachjustieren der einzelnen Operationen an beiden Werkzeugmaschinen.

Tabelle V
ZEITAUFWAND FÜR DAS NACHJUSTIEREN ALLER OPERATIONEN AN DER *Fanuc Robodrill α-D21MiB5* UND *Doosan Puma 2600SY II*.

Operation	*Robodrill α-D21MiB5* t / h	*Puma 2600SY II* t / h
Rohteil aufnehmen	0.25	-
Rohteil bestücken	0.10	-
Fertigteil entnehmen	0.10	0.10
Bauteil reinigen	0.00	0.00
Fertigteil ablegen	0.10	0.25
Späneerkennung Bauteil	3.00	3.00
Späneerkennung Spanner	0.00	0.00
Automatische Tür öffnen	0.00	0.00
Automatische Tür schließen	0.00	0.00
CNC Start	0.00	0.00
Summe	3.55	3.35

Den größten Aufwand bildet erneut das Entwickeln von zwei Kamerajobs für die Späneerkennung am Bauteil. Bei allen neuen Bauteilen müssen zwnagsläufig neue Kamerajobs entwickelt werden, da die vorhandene Bauteilkontur für die Auswertung der Späneerkennung relevant ist. Dieser hohe Aufwand entfällt, wenn ein bereits bestehenden Bauteilen erneut produziert wird. Die aufgezeichneten Daten zeigen, dass der Umrüstprozess bei allen Werkzeugmaschinen in weniger als einem halben Arbeitstag durchgeführt werden kann.

C. Evaluierung der Prozesssicherheit

Einer der wohl wichtigsten Faktoren eines jeden Automatisierungssystems ist die vorhandene Prozesssicherheit der Anlage. Eine prozesssichere Automatisierung zeichnet sich durch eine konstante und zuverlässige Produktion von Bauteilen mit wenig Fehlerfällen aus. Der entwickelte Funktionsprototyp wird anhand einer kurzen Testproduktion von 60 Bauteilen auf seine vorhandene Prozesssicherheit überprüft, sodass die häufigsten Fehlerquellen ermittelt werden können. Die Testproduktion erfolgt an der Werkzeugmaschine *Doosan Puma 2600SY II*. Alle auftretenden Fehler werden kategorisiert und in Abbildung 9 grafisch aufgezeigt.

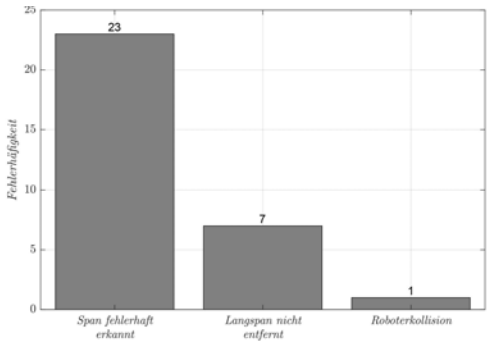


Abbildung 9. Übersicht über die Fehlerhäufigkeit während der Testproduktion von 60 Bauteilen.

Jede Fehldetektion eines Spans wird durch die vorhandene Emulsion hervorgerufen. Dabei wird das Reinigungsprogramm fälschlicherweise gestartet, wodurch ein erhöhter Zeitaufwand entsteht. Dieses Problem kann durch zusätzliches Bildmaterial an der Kamera behoben werden. Die Roboterkollision wird von der vorhandenen Schleppkette verursacht, welche sich nach mehrfacher Belastung lockert. Dieses mechanische Problem kann durch die Verstärkung aller Befestigungspunkte gelöst werden. Das größte Problem bezogen auf die Prozesssicherheit bildet das Auftreten von Langspänen. Trotz durchgeführter Nachjustierungen am Reinigungszyklus können Langspäne nicht zuverlässig entfernt werden. Diese Späneart tritt bei Drehmaschinen häufig auf, wodurch in unregelmäßigen Abständen mit einer Unterbrechung der laufenden Produktion gerechnet werden muss. Aufgrund dieser Tatsache ist mit dem vorhandenen Aufbau keine prozesssichere Produktion an einer Drehmaschine realisierbar. Dieses Problem tritt bei Fräszentren nicht auf, da dort in der Bearbeitung nur kurzbrechende Späne entstehen. Dadurch ist die Prozesssicherheit des Systems an Fräszentren wie der *Fanuc Robodrill* α-*D21MiB5* wesentlich höher als an einer Drehmaschine.

V. Conclusio

Das entwickelte Konzept besitzt ein hohes Potential KMU's eine geeignete flexible Automatisierungslösung für verschiedenste Produktionsaufgaben bieten zu können. Der modulare Aufbau und die objektorientierte Programmierung ermöglichen dabei eine hohe technische Flexibilität, wodurch sich jegliches System an die vorhandenen Anforderungen anpassen kann. Dank der realisierten Moderator-Teilnehmer Struktur ist es möglich ein allgemeines Ansteuerungsprogramm zu entwickeln, wodurch sich anfallende Programmierarbeiten bei Änderungen verringern. Aufgrund der Tatsache, dass alle Funktionalitäten der Anlage auf der Logikeinheit der Teilnehmer implementiert sind, kann über Methodenaufrufe und standardisierte Schnittstellen jede Funktion zentral gestartet und koordiniert werden. Die frei parametrisierbare Ablaufkette ermöglicht es, den genauen Produktionsprozess individuell an die vorhandene Werkzeugmaschine anzupassen. Die erhobenen Messdaten weisen auf ein funktionsfähiges System und eine gute Performance hin, jedoch müssen für ein serienreifes Automatisierungssystem alle Mängel der Prozesssicherheit behoben werden.

Danksagung

An dieser Stelle möchte sich der Autor bei seinen Betreuern Benjamin Massow und Armin Lechner für die Ermöglichung des Projekts bedanken. Ein weiterer Dank geht an die Kommilitonen Andreas Mischi, Sebastian Erler, Laurenz Stöckl und Lukas Dinter, die maßgeblich an der Umsetzung des Projektes beteiligt sind.

Literatur

[1] T. Ludwig, C. Kotthaus, M. Stein, H. Durt, C. Kurz, J. Wenz, T. Doublet, M. Becker, V. Pipek und V. Wulf, "Arbeiten im Mittelstand 4.0 – KMU im Spannungsfeld des digitalen Wandels," *HDM*, pp. 71–86, 2016.

[2] E. Bogner, U. Löwen und J. Fanke, "Bedeutung der zukünftigen Produktion kundenindividueller Produkte in Losgröße 1," in *Interdisziplinäre Perspektiven zur Zukunft der Wertschöpfung*, 1st ed. Abraham-Lincoln-Str. 46, 65189 Wiesbaden, Germany: Springer Fachmedien Wiesbaden GmbH, 2018, ch. Teil 1, pp. 63–75.

[3] A. Lechner, "Ausarbeitung eines Geschäftsmodells zur flexiblen Produktionsautomatisierung mit mobiler Datenintegration in fertigungsorientierten mittelständischen Unternehmen," Masterarbeit, Hamburger Fern-Hochschule, 15. Nov. 2019.

[4] M. Thaler, "Entwicklung eines modularen Cyber-Physical System zur Automatisierung von rekonfigurierbaren Kleinserienfertigungen," Masterarbeit, MCI | Die Unternehmerische Hochschule®, 29. Juli. 2020.

[5] L. Dinter, "Umsetzung einer modularen sowie mobilen Roboterplattform zur flexiblen Automation von Bearbeitungsmaschinen in Produktionsunternehmen," Bachelorarbeit, MCI | Die Unternehmerische Hochschule®, 24. September. 2021.

[6] S. Erler, "Sicherheitstechnische Betrachtung von modularen, kollaborativen Automationssystemen und Entwicklung eines Leitfadens für Anlagen Änderungsprozesse unter Erhalt der CE-Zertifizierung," Bachelorarbeit, MCI | Die Unternehmerische Hochschule®, 31. Juli. 2021.

[7] L. Stöckl, "Sicherheitstechnische Betrachtung von modularen, kollaborativen Automationssystemen und Konzeptionierung von mobilen Umsetzungsvarianten," Bachelorarbeit, MCI | Die Unternehmerische Hochschule®, 31. Juli. 2021.

[8] A. Mischi, "Entwicklung eines Detektions- und Reinigungskonzepts von Bearbeitungsresten im Automatisierungsprozess von Bearbeitungsmaschinen," Bachelorarbeit, MCI | Die Unternehmerische Hochschule®, 24. September. 2021.

[9] DIN EN 60204-1 VDE 0113-1:2019-06, "Sicherheit von Maschinen – Elektrische Ausrüstung von Maschinen - Teil 1: Allgemeine Anforderungen," (IEC 60204-1:2016, modifiziert).

[10] DIN EN 61439-1 VDE 0660-600-1:2012-06, "Niederspannungs-Schaltgerätekombinationen - Teil 1: Allgemeine Festlegungen," (IEC 61439-1:2011).

[11] DIN EN 61439-2 VDE 0660-600-2:2012-06, "Niederspannungs-Schaltgerätekombinationen - Teil 2: Energie-Schaltgerätekombinationen," (IEC 61439-2:2011).

[12] DIN EN 61355-1 VDE 0040-3:2009-03, "Klassifikation und Kennzeichnung von Dokumenten für Anlagen, Systeme und Ausrüstungen - Teil 1: Regeln und Tabellen zur Klassifikation," (IEC 61355-1:2008).

[13] DIN EN IEC 81346-2:2020-10, "Industrielle Systeme, Anlagen und Ausrüstungen und Industrieprodukte - Strukturierungsprinzipien und Referenzkennzeichnung - Teil 2: Klassifizierung von Objekten und Kennbuchstaben für Klassen," (IEC 81346-2:2019).

Marcel Stifter studiert Mechatronik am MCI in Innsbruck/Österreich. Seine Hauptkompetenzen liegen im Bereich Elektrotechnik und Programmierung von Hardwarekomponenten. Neben dem Studium arbeitet er an der Entwicklung von Elektromotoren im Bereich Automotive in der Firma GKN Driveline in Bruneck/Italien.

Automated Oxygen Supply System For Avalanche Victims

Armin Stockhammer and Bernhard Hollaus (supervisor)

***Abstract*—The steadily increasing demand for outdoor activities leads to a higher number of casualties. Especially in DACH countries ski touring and skiing off-piste is becoming more popular and entails avalanche accidents.**
Depending on the snow conditions and the personal fitness, completely buried avalanche victims can breathe up to $30\,\mathrm{min}$ until they run out of air. Rescue operations often take several hours to find and dig out the buried skiers. The goal is to make such dangerous sports safer and increase the chances of surviving an avalanche accident. The thesis focuses on a prototype that supplies fresh air without active interaction by the carrier. Therefore the oxygen of blown-up airbag-systems is used and controlled by an embedded system (ES). Depending on the measured ambient air quality the ES transports fresh air out of the airbags towards the victim's face. Due to several continuous pressure and gas monitorings the pneumatic structure is efficiently controlled for a maximum yield of oxygen and hence, an expanded supplying time. By using the Bluetooth Low Energy (BLE) technology the current system status is broadcasted to nearby Internet-of-Things (IoT) devices. The provided information is important to rescue teams or comrades and are tracked live on the developed window application.
The test results have proved that the fresh air supplied by the prototype slows down the rise of the pollution level in the air pocket thereby increasing the victims' chances of not dying from suffocation. Hence the prototype is a valuable safety equipment and an eligible candidate for the product market.

***Index Terms*—oxygen supply control, emergency gadget, breath-VOC monitoring, index of air quality, Bluetooth Low Energy, increased survival chance**

I. Introduction

TODAY, back country skiing and ski touring off-piste are trending sports and hence contributing to a higher accident rate in the mountains. Major consequences arises for the rescue teams who are putting themselves in danger each time. According to a alpine accident statistics of the Austrian Alpine Safety Board [1] 199 avalanche incidents were noticed in total in Winter season 2020/2021 whereof 14 people lost their lives. Although many protective gears were developed in the last decades such as airbags, the primary cause of death from snow avalanches is asphyxia [2]. Within 35 minutes approximately $70\,\%$ of completely buried victims die due to a lack of oxygen [3],[4],[5].
Steadily new developed technologies ensure improvement and optimization on current products resulting in a contest at the marketplace. A shovel, a probe and an avalanche transceiver are basic equipment for back country skiers today. These tools are helpful in case of an avalanche accident but they do not support the athletes in the actual event. Based on the principle of rescue emergency systems can be categorized into [6]:

a) localization of buried people without carrying a transceiver
b) rapid localization of buried victims, already by fellows in order to lower the endangerment for a rescue team
c) prevention of being buried by an avalanche
d) extended life-time while being buried

The first group contains two systems that yield the lowest chance of survival. One of them is a specially trained avalanche dog that needs to be transported to the location in the first place is deployed. The other system is the *RECCO* rescue technology that is able to locate buried or lost people outdoor. It's a two-part system, featuring an active detector, carried by the rescuer, and a passive reflector, carried by the user. The detector emits a directional radar signal and is echoed by the reflectors. Depending on the strength of the returned signal, the rescuer can locate the victim. The *RECCO* reflectors are non-powered and are commonly integrated in different outdoor clothes, which increases the total reflection area in order to be found more likely.
Within group B buried persons are located by different gadgets such as an avalanche beacon. Together with a shovel and a probe, they offer a standard equipment for back-country skiers. However, the requirement is that both rescuer and victim carry a receiver/transceiver being switched on to the operating mode *SEARCH* and *SEND* respectively. The principle is that a signal is broadcasted on a carrier frequency of $457\,\mathrm{kHz}$ $\pm 80\,\mathrm{Hz}$ by one or more antennas. Three antennas facilitate the best distribution and highest probability of being found given that they are arranged like an orthogonal coordinate system. Also the specific signal frequency is a harmonized European standard defined in ETSI EN 300 718-1 [7] which provides the longest range through snow.
Another gadget called *AVALANCHE BALL* was invented by the Tyrolean duo D. Venier and H. Fournier in 1996. The system consists of the buoyancy body (ball), a $6\,\mathrm{m}$ connection leash and its mounting and storage bag. The lifesaver must be triggered manually by the user in the event of being swept by an avalanche. The integrated spring blows up the ball and will remain on top of the avalanche even when a second avalanche comes across. Its color and connection leash will lead the rescuer or helping comrade to the victim immediately.
Gadgets of Group C are preventive tools such as the avalanche airbags. These inventions are based on the principle of inverse segregation in which objects whose surface area is larger than

A. Stockhammer studies Mechatronics and Smart Technologies at the MCI, Innsbruck, Austria, e-mail: sa4113@mci4me.at.
Paper submitted on 30. September, 2021; revised on 30. September, 2021

surrounding snow crystals get forced to the surface. The body's volume is increased by the inflated airbag and adds buoyancy to keep the body on the surface of a moving mass of snow [8].A common airbag has a volume of 150 l up to 170 l like the *ABS TwinBag* system used in the ABS A.LIGHT Avalanche Airbag. Avalanche airbags are a valuable safety device but don't guarantee survival. They barely protect against mechanical injuries and require an active triggering at the time of an avalanche descent. Their effectiveness is evaluated in a clinical paper [9] where different cases, factors and circumstances are considered and included in the calculation. The overall adjusted mortality is 11 % (95 % confidence interval from 6 % to 16 %) for victims with inflated airbags and 22 % (95 % confidence interval from 15 % to 29 %) for victims with no or non-inflated airbags. According to that paper [9] these gadgets can lower the mortality by directly reducing the grade of burial which is the most important factor for survival.
All previously described emergency devices are supportive in particular situations but can't expand the survival time in a critical burial position. Based on the previous elucidated mortality rate and the consideration of all different types of avalanche involvements, the chart in figure 1 is generated.

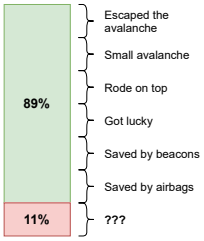


Fig. 1. Distribution of avalanche cases

Whereas some people die from a hit by a big massive object or due to falling from great height the primary cause of death through snow is asphyxia [2]. One product facing the danger by suffocation is the *Emergency breathing device* [10] which was already patented by T. Crowley and further refined and commercialized as *AvaLung* by the company *Black Diamond* [11]. The gadget is a wearable device with a mouthpiece, tubing and valves in order to control and filter the outgoing/incoming air. It enables the victim to breathe under the snow for the purpose of gaining valuable time for being rescued. According to several feed-backs from the field this emergency equipment is described as unpleasant and impracticable, which urged the company to take it off the market.
Another different patented product is the *protective helmet for air extraction from snow* [12]. The apparatus and system contain air intake cavities disposed on the outer surface of the protective helmet and on the outer backpack structure in order to feed ambient air to the victim by a mouthpiece.
All these products provide a solution for avalanche buried people but require that the skier actively intervene by putting the mouthpiece in the mouth. Therefore, it would be of interest to develop a system that feeds extra oxygen to the victims at the time of low oxygen supply. The additional air is extracted from triggered airbag backpacks. The embedded system monitors the breathing gases and depending on the current air quality fresh air is dispensed by the system. All data is broadcasted via Bluetooth Low Energy (BLE) to potentially receiving systems nearby. The prototype answers the leading question whether the victims' survival chance can be increased due to an oxygen supply or not.

II. Methods

For the development of an automated oxygen supply system the target group for the product has to be specified. The advantages or disadvantages of an independent system compared to a modular gadget which is highly applicable on current systems are visualized in table I. It has to be noticed that these 5 listed criteria/areas are based on a project management concept called *TELOS* and considers the feasibility of a project. Additionally, the areas are particularly considered with regard to the terms of a master thesis.

TABLE I
Advantages and disadvantages of an independent and of an modular applicable system.

Criteria:	Independent system	Modular applicable system
Technical feasibility	-	+
Economic aspects	+	+
Legal aspects	+	+
Operational	-	+
Scheduling	-	+

Based on the positive aspects above and the provided ABS airbag-systems the idea of a modular applicable system is selected. The basic concept of the prototype is depicted in figure 2 containing both airbags, the cartridge filled with compressed air, the embedded system and the air supply tube.

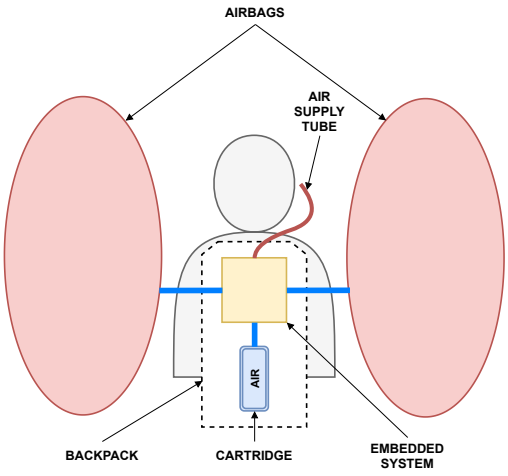


Fig. 2. Basic concept of the prototype

The embedded system is the main unit of the gadget and is responsible for distributing and controlling the air. Hence the ES is disposed between the airbags and the source of the air. In case the ABS-system is triggered and the airbags are blown-up the pneumatic structure in between ensures a closed system. As a consequence, the trapped air can easily be controlled and forwarded to the user's face through the supply tube.

A. Hardware - Implementation

Hardware is composited in an embedded system and is categorized into mechanical components and electrical components. The mechanical construction is designed in order to fulfill following requirements:

1) Applicability to different systems on the market
2) Processing up to $p_C = 300\,\text{bar}$ of compressed gas
3) Environmental conditions (e.g. low temperatures and humidity)
4) Reliability

The first requirement is ensured by using an adapter which acts as a link between the specific ABS gas outlet component and the embedded system. By providing various adapters for different systems the oxygen supply system can be integrated easily.
For safety reasons the system is developed to withstand the maximum occurring gas pressure of $p_C = 300\,\text{bar}$. According to airbag manufacturer *ABS Protection GmbH* their cartridges release a pressure of $p_C = 300\,\text{bar}$ leading to a single reservoir inlet pressure of around $p_{RSV} = 160\,\text{bar}$ in the first second after the cartridge has been tapped. After about $3\,\text{s}$ the airbags are fully blown-up. The law of Boyle-Mariotte - $p_1 \cdot V_1 = p_2 \cdot V_2$ -, in which the absolute temperature T and the amount of gas substance n are assumed as constants, enables the calculation of the airbag pressure after being blown-up. The cartridge has a pressure of $p_C = 300\,\text{bar}$ within a volume $V_C = 0.7 \times 10^{-3}\,\text{m}^3$ whereas the total airbag volume of the *ABS TwinBag System* is $V_{AB} = 170 \times 10^{-3}\,\text{m}^3$. Considering these values by using the law of Boyle-Mariotte it results in equation 1.

$$p_{AB} = \frac{p_C \cdot V_C}{V_{AB}} = \frac{300\,\text{bar} \cdot 0.7 \times 10^{-3}\,\text{m}^3}{170 \times 10^{-3}\,\text{m}^3} = 1.24\,\text{bar} \quad (1)$$

Furthermore all components are selected to resist all different kinds of ambient conditions like temperature and humidity. Hence the function of the components has to be ensured within temperature range from about $-20\,°\text{C}$ up to $30\,°\text{C}$.
Having regarded all requirements the resulting pneumatic build-up concept is depicted in figure 3 focusing on the high pressure scope.

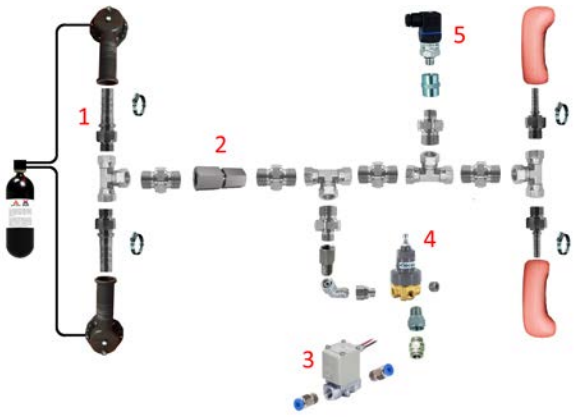


Fig. 3. Pneumatic build-up concept - high pressure scope

As explained in figure 2 the ES works as bridge unit between the two outlets of the ABS system and the two airbags. Pneumatic adapters (no. 1 in figure 3) are used to connect the ABS specific *Venturi units*. The adapter solution allows the user to connect to their airbag system independently from the manufacturer. The hose fitting is a standardized article which can be selected in different sizes.
The axial high-pressure check valve (no.2 in figure 3) prevents the passed air from flowing back and to remain in the closed system respectively.
In order to extract air from the closed system another valve (no. 3 in figure 3) is integrated. This solenoid valve is electrically controllable and hence opening and closing the enclosed air. Additionally, a high pressure regulator (no. 4 in figure 3) is arranged prior to the magnetic valve. This component provides a lower output pressure allowing the use of a low pressure magnetic valve, which in turn requires low control current.
The second pneumatic build-up focuses on the low pressure controlling and the corresponding embedded system. This concept is depicted in figure 4.

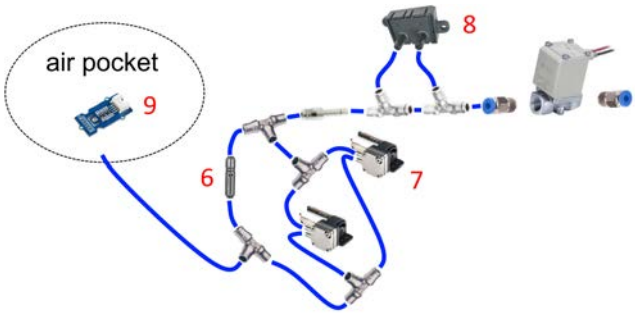


Fig. 4. Pneumatic build-up concept - low pressure scope

The low pressure unidirectional valve (no. 6 in figure 4) is connected parallelly to two electrical air pumps (no. 7 in figure 4). This has to be implemented in order to protect the pumps against high pressure damage and to prevent transported air from flowing back.
Based on the idea that the gas sensor (no. 9 in figure 4) monitors the compounds in the air pocket of a buried victim, a special fixture is required. The air supply housing is located at the carrier's chest as depicted in figure 5.

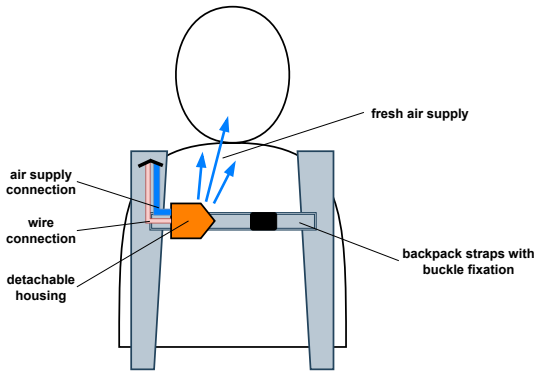


Fig. 5. Air supply unit concept

The position of the air supply unit is off-centered to the sagittal plane and close to the user's face. The closeness is essential in order to supply fresh air without additional active actions by the victim. Within the housing the gas sensor, the wire and pneumatic connectors are mounted.
The implementation of the electronic system requires several components which are assembled on a prototype PCB and controlled by a micro-controller unit (MCU). The logical build-up is categorized by color and illustrated in figure 6.

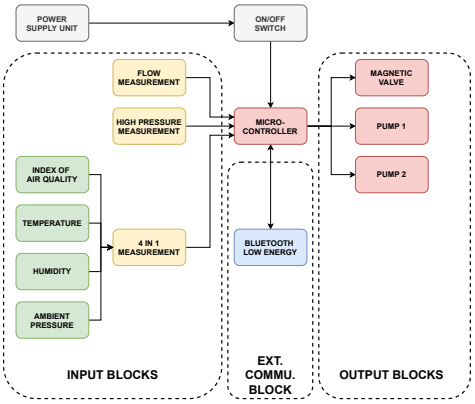


Fig. 6. Logical buildup of the electronic system

The entire power supply of the embedded system is electrically separable with a ON/OFF switch allowing the user to disconnect when not in use and to protect jack and socket against unnecessary connection cycles. All red colored blocks in figure 6 are components controlled by the MCU and represent the MCU-outputs. As derived from the figure above, two pumps are implemented due to their maximum flow rate of $Q_{pump} = 380 \times 10^{-3}\,\mathrm{m^3/min}$ only. The three yellow blocks define the MCU-inputs indicated by arrows pointing towards the micro-controller. A high pressure measurement (no. 5 in figure 3) is necessary to recognize the airbags being blown-up. The additional airflow sensor (no. 8 in figure 3) detects if a airflow still exists and hence activating the pumps in the other case. The green blocks are displayed to emphasize the measured parameters recorded by a single sensor. Finally, the blue *Bluetooth Low Energy* (BLE) device is used to communicate with external *Internet of Things* (IoT) devices.
Combining all corresponding component circuits in the ALTIUM Designer it yields the developed prototype of the printed circuit board (PCB). The PCB consists of two conductive layers bottom and top only. The circuit board assembly (CBA) of the developed PCB design is depicted in figure 7.

Fig. 7. Extension board for MCU Arduino Mega 2560

B. Software - Implementation

The control code is programmed in Arduino where the procedure - represented by the `loop`-function - is visualized in figure 8.

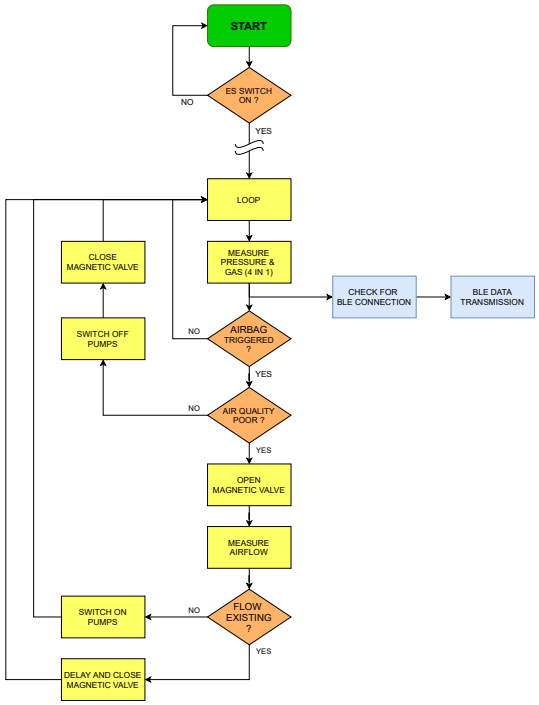


Fig. 8. Control procedure of the embedded system

The main Arduino functions are enumerated and described as following:

- `FS_High_read`-function
  This method reads input values from the high pressure transmitter (no. 5 in figure 4) and converts them to pressure values in bar. The real value in the unit *bar* is obtained by using equation 2.

$$FS_{H\_bar} = \frac{(FS_{H\_value} - FS_{H\_init}) \cdot 250}{FS_{H\_scale\_factor} \cdot 4} \tag{2}$$

  In case the measured value is higher than $15\,\mathrm{bar}$ the boolean variable `airbag_triggered` is set true. $15\,\mathrm{bar}$ is selected as a boundary value because of the sensors inaccuracy of $\pm 1\,\%$ and the signal noise. After the boolean is set as true a time delay of $3\,\mathrm{s}$ follows. This setting is required to prevent the ES from starting the air controlling until the airbags are fully blown-up - which takes around $3\,\mathrm{s}$ according to the manufacturer ABS.
- `FS_Low_read`-function
  Equal to the high pressure sensor the interpreted analog input voltage of the flow sensor is converted into an applied flow value FS_{L_FA} known as a SCCM value by using equation 3. The formula is derived from the sensor's datasheet [13].

$$FS_{L\_FA} = \frac{FS_{L\_F\_fs} \cdot \left(\frac{FS_{L\_value}}{FS_{L\_V\_s}}) - 0.5 \right)}{0.4} \tag{3}$$

FS_{L_value} is the value of the interpreted analog input voltage, $FS_{L\_F\_fs}$ the fullscale value of the selected sensor type and $FS_{L\_V\_s}$ the sensor's supply voltage.

- `GS_read`-function
  The `GS_read`-function reads all parameters from the gas sensor. As the sensor mode is preset to the lower power mode at the beginning of the program via the variable $bsec\_config\_iaq$ data values are gained every third second. A time-stamp variable helps to record the time for later analysis. The received pressure, temperature, humidity and Index of Air Quality (IAQ) values are already compensated via the specific BSEC algorithm. However, the pressure value is returned in Pascal unit which requires a conversion in order to gain the atmospheric height value in meter with respect to the sea level. This is approached by using the barometric formula where the height is solved through the international standard atmosphere according to equation 4.

$$GS_{ht} = \left(\frac{288.15\,\mathrm{K}}{0.0065\,\mathrm{K\,m^{-1}}}\right) \cdot \left(1 - \left(\frac{pressure(h)}{1013.25\,\mathrm{hPa}}\right)^{\frac{1}{5.255}}\right) \tag{4}$$

- `BLE_input_check`-function
  Preconditioned that a BLE device is already connected this function checks for incoming requests for starting or stopping the data transmission.
- `BLE_data_transmission`-function
  In case the BLE request contains a `start`-command the current parameters of each sensor are sampled by the `sampling_for_BLE`-function. Once the first data is sent, the next data will be sent only if the old data has already been sent.

Like previously mentioned the entire data are sent to potentially receiving devices like rescue-teams or comrades. Therefore a proper application is required to interpret and display the transmitted information. *Qt* is a so-called *widget toolkit* for designing graphical user interfaces (GUI) and cross-platform applications which run on various software and hardware platforms. By using the GUI of Visual Studio with Qt toolkit the following application is developed for the user, see figure 9.

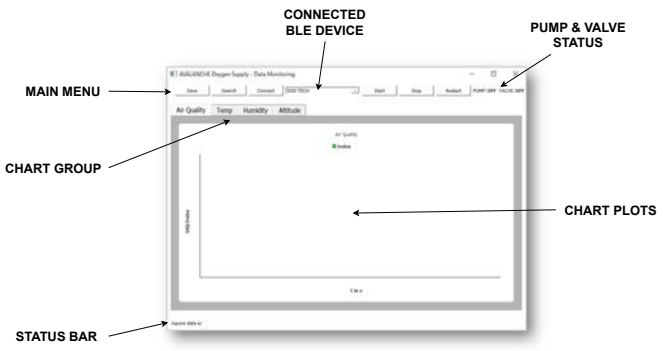


Fig. 9. Window application of the user

In order to focus on different aspects of the prototype, two test conditions are defined as following.

- Test scenario 1
  In the first test all components are assembled as intended except for the connection to the ABS airbag system. The system with the air cartridge is replaced by an electric air compressor providing the pressure to blow up the airbags. The main goal of this test is to check the air tightness of the system and the functional verification of the embedded system. The compressor is not able to activate the air control loop as the bounding pressure value of $15\,\mathrm{bar}$ can't be obtained. Hence the code is slightly adapted and directly activated after the power switch is toggled to *ON*.
  Furthermore, in order to accomplish an approximated simulation of a buried victim a four liter sized plastic box is representing the air pocket. This size is selected based on a scientific paper [5] where the effects of snow properties on humans breathing into an artificial air pocket are investigated. As natural air pockets in the snow are not completely sealed 36 drilled holes at $2\,\mathrm{mm}$ in diameter are arranged over the surface box, see figure 10.

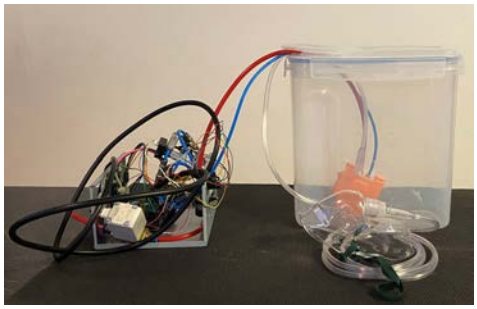


Fig. 10. Test build-up 1

  These holes enable the test user to breathe ambient air or emit breath-VOC into the surrounding equal to the snow characteristics. The test person uses a breathing mask thereby. As a precaution the IAQ boundary level for activating air supply is set to 150 (0-excellent, >351-extremely polluted) where no significant irritation is possible (see IAQ table in gas sensor's datasheet [14]).
- Test scenario 2
  The second test checks the pneumatic structure - see also figure 3 - by using the integrated cartridge from the ABS Airbag-system. The only available carbon cartridge with electrical triggering is filled with Nitrogen-gas and therefore can not be used with test persons involved. The focus of this test is to check if the embedded system starts controlling after the airbags are blown-up by the cartridge.

III. Results

After completing the development of the hardware (electronic and pneumatic part) as well as the software (Arduino controlling, Qt BLE application) all single functions of the

prototype are tested. This assumes that everything has been assembled accurately.
First of all the designed CBA is assembled accordingly. Due to time constraints a stripboard is used to build up the circuitry with its components. All electrical components are connected to the extension board and plugged onto the MCU at the very left as depicted in figure 11.

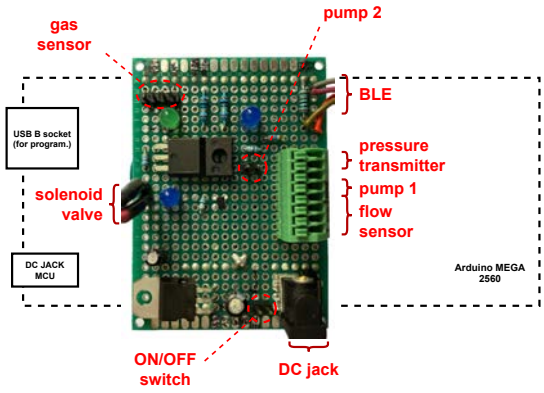


Fig. 11. Connectors on the extension board

The air supply housing is produced with the *Anycubic i3-mega* 3D-printer leading to the following results in figure 12.

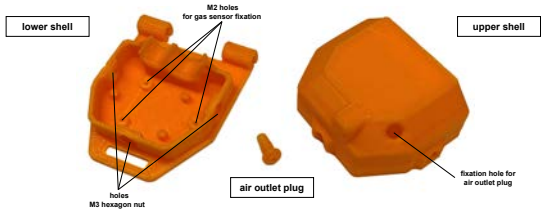


Fig. 12. Air supply unit - 3D prints

As shown in figure 13, the air supply unit is able to be mounted at the intended position but can be adjusted in height and combined with the buckle strip.

Fig. 13. Mounted air supply unit

The entire embedded system working with low pressure only is mounted into a standard PC power supply unit housing where different holes are used to fixate all components - see figure 14.

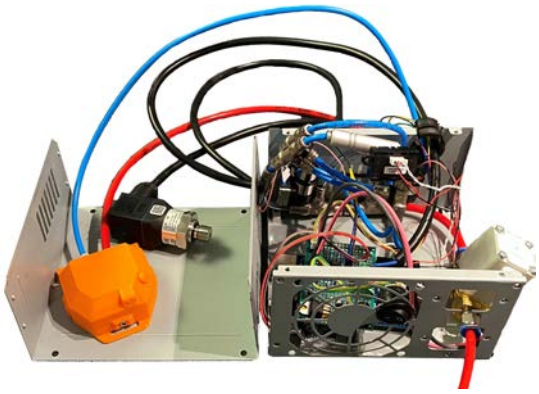


Fig. 14. Embedded sytem with opened housing

In order to perform the first test according to scenario 1 a proper power supply unit with $V_{CC} = 12\,\mathrm{V}$ is plugged in on the extension board as well as USB-B cable to internal monitor the serial prints. During a pre-test it has been recognized that the IAQ seems to remain at a certain index of around 110. Thus it's assumed that the drilled holes are too big and can't represent the porosity of real snow. Like a deviation to the defined test-plan, the holes are sealed by a PTFE tape. The test results are depicted in the following figures, displaying the IAQ in figure 15 and the true altitude in figure 16.

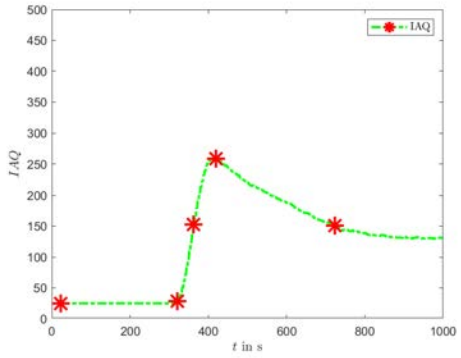


Fig. 15. Test results of scenario 1 - IAQ

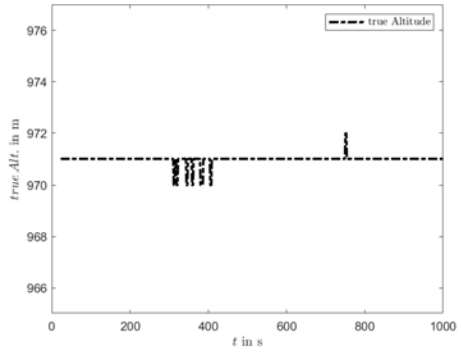


Fig. 16. Test results of scenario 1 - true altitude

In the IAQ chart five specific data points are highlighted. The first one indicates the first measured data, which is slightly

time-shifted as the BLE connection is started a short moment after the embedded system has been switched on. Marker 2 depicts the event when the gas sensor has finished its initialization time of three minutes. At this time the breathing mask is positioned in the box which is being filled with exhaled air. After further 50 sec the boundary value of 150 IAQ is reached, which triggers the control cycle and changes the displayed status of the valve and the pumps. As a consequence of the transported fresh air the rapid rise of the index is decreasing. Since a critical well-being value is reached at over 250 IAQ the test is discontinued by pulling the breathing mask out of the box. The quite linear slope between marking point four and five represents the air supplying feature leading to steadily improved air quality within the box. The difference is more visible by comparing it to the IAQ behavior after the system switches off at the last marker.
The true altitude in figure 16 depicts a constant value as expected. A few outliers are caused by noise and rounding errors thereby showing the relation of the height calculation which depends on pressure, humidity and ambient temperature.
Moreover the monitored results of the temperature and the corresponding humidity are shown in figure 17 and figure 18 respectively.

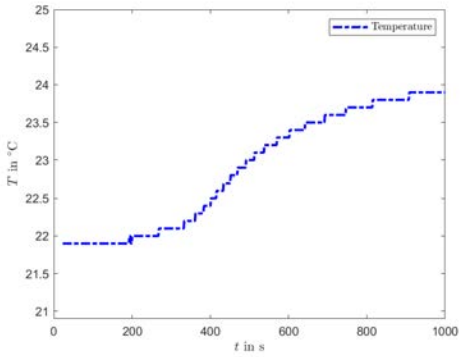


Fig. 17. Test results of scenario 1 - temperature

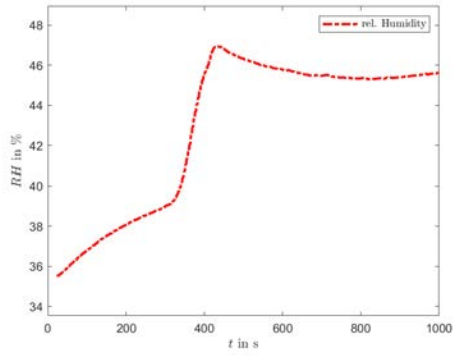


Fig. 18. Test results of scenario 1 - humidity

The graphs above show a noticeable value increase after the initialization process, which is equal to the start of breathing with the mask. After breathing has been discontinued the monitored data shows a decreasing tendency of the parameters.

During the activated control process the GUI signals the status of the valve and the pump at the top right corner as shown in figure 19.

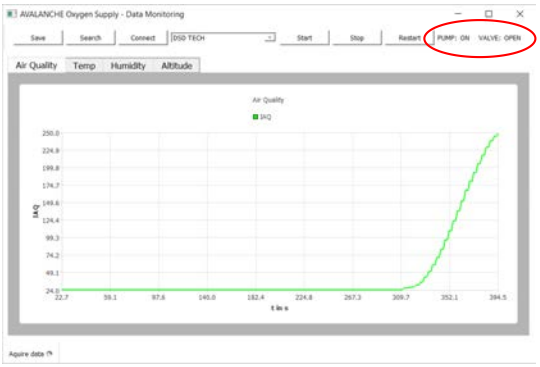


Fig. 19. Live monitoring by Qt application - IAQ

These results prove the intended functionality of the low pressure structure of the embedded system.
Concerning the second test, all components of the high and low pressure structure are assembled leading to the final test build-up in figure 20.

Fig. 20. Test build-up 2

Once the embedded system has been power supplied and connected to the laptop's Qt-application via BLE, the T-handle bar is pulled. The electrical cartridge tapping is recognized but no air flow has been following. This means that the provided carbon test cartridge is not filled. As no further applicable cartridge is available, the test conditions are redefined. In order to prove tightness and trigger recognition by the high pressure measurement transmitter, the airbags are manually blown-up by an air compressor. Air is directly blown into the high-pressure unidirectional valve because this component is easily accessible and mandatory to prevent air leakage. As only a maximum pressure of $p_{comp.} = 8\,\mathrm{bar}$ can be attained, the boundary detection level for the high pressure transmitter is decreased to $5\,\mathrm{bar}$ in the Arduino code. The build-up of the post-test is depicted in figure 21.

Fig. 21. Test build-up 2 - post-test

After the filling process the embedded system waits until the gas sensor initialization has been finished. Meanwhile, deflation of the airbags is not perceptible which indicates that no leakage is given in the high pressure structure. Finally, by exhaling into the artificial air pocket, the system is urged to start the control process due to bad air quality. The status LED of the valve lights up and thus signalizes that the high pressure transmitter has interpreted the air flow correctly.

IV. Conclusion

The oxygen supply system proves to be a valuable avalanche protection gear. Based on those results it can be assumed that the victims' chances of not dying from suffocation are thereby reduced. Despite achieving all defined goals, some improvements will necessarily have to be performed in order to provide a marketable product. One drawback is seen by the total mass of the device. Due to the design for high pressures, the components have to be made out of steel or other robust materials. However, such parts can be integrated into a new airbag system where the total mass can be decreased significantly and thereby lead to an efficient solution. Further tests under real snow conditions could validate the usability of the air supply unit which is mounted on the backpack strap. This will also reveal the maximum transmission range of a buried victim. A future prospect is to expand the time of supply by using pure or highly concentrated oxygen air mixture. Hence less amount of air is required by the avalanche victim yielding to a longer timespan.
All in all, the achieved goals can be used as a starting point for further research projects, in the best case an optimized prototype will be leading to a commercialized product.

Acknowledgment

First the author would like to thank his thesis supervisor Bernhard Hollaus, PhD. of the Management Center Innsbruck(MCI), who has supported him throughout the research and development phases. The door to his office was always open whenever the author ran into a trouble spot or had questions regarding a specific issue. As this is not his only thesis to be supervised the author really appreciates his quick responses to questions, mails or meetings.

The author would also like to acknowledge a.o. Univ.-Prof. MMag.Dr. Alexander Hörbst, Bakk.techn. course director of *Medical & Health Technologies* at the MCI, who has been an additional founder of the overall topic. He has connections to the manufacturer ABS Protection GmbH who provided several ABS airbags for research projects. At this point, the author would also like to thank ABS Protection GmbH for the backpacks.

Finally, the author want to thank the Management Center Innsbruck itself for its financial support in order to make this prototype development feasible.

References

[1] J. Zauner, "Alpinunfälle - Bericht Winter 2020/2021," Presseaussendung, Österreichisches Kuratorium für Alpine Sicherheit, April 2021, https://www.alpinesicherheit.at/data/docs/2021/Presseaussendung_April%202021_OeKAS_Winter%202020-21%20Alpinunfallzahlen.pdf.

[2] H. Brugger, J. H. Etter, J. Boyd, and M. Falk, "Causes of death from avalanche," *Wilderness Environ Med*, vol. 20, pp. 93–96, 2009.

[3] H. Brugger, B. Durrer, L. Adler-Kastner, M. Falk, and F. Tschirky, "Field management of avalanche victims," *Resuscitation*, vol. 51, pp. 7–15, 2001.

[4] H. Brugger, L. Adler-Kastner, and M. Falk, "Avalanche survival chances," *Resuscitation*, vol. 368, p. 21, 1994.

[5] G. Strapazzon, P. Paal, J. Schweizer, M. Falk, B. Reuter, K. Schenk, H. Gatterer, K. Grasegger, T. Dal Cappello, S. Malacrida, L. Riess, and H. Brugger, "Effects of snow properties on humans breathing into an artificial air pocket – an experimental field study," *Scientific Reports*, vol. 7, no. 1, p. 17675, Dec 2017. [Online]. Available: https://doi.org/10.1038/s41598-017-17960-4

[6] H. J. Etter, J. Schweizer, and T. Stucki, "Lawinennotfallsysteme im Vergleich: Nicht ohne mein LVS," *Alpen*, no. 85, pp. 24–29, 2009.

[7] *Avalanche Beacons operating at 457 kHz; Transmitter-receiver systems; Part 1: Harmonised Standard for access to radio spectrum*, ETSI Std. EN 300 718-1, Rev. 2.2.1, 2021-06.

[8] J. Blevins, "Colorado avalanche danger worst in 30 years — inside and outside of ski areas, experts say," Newspaper, The Denver Post, February 2012, https://www.denverpost.com/2012/02/19/colorado-avalanche-danger-worst-in-30-years-inside-and-outside-of-ski-areas-experts-say/.

[9] H. Haegeli, M. Falk, E. Procter, B.Zweifel, F.Jarry, S. Logan, K.Kronholm, M. Biskupic, and H. Brugger, "The effectiveness of avalanche airbags," *Resuscitation, Elsevier*, no. 85, pp. 1197–1203, 2014.

[10] C. Hirschfield, "The Avalung: A Look at the Backcountry's Newest Player," *Backcountry Magazine*, p. 42, Dec. 1999.

[11] N. Beidleman, J. Margid, T. Crowley, and R. Siberell, "Emergency breathing device," Europe Patent Application EP0 998 959A1, May 10, 2000.

[12] S. A. Ogilvie, "Protective helmet for air extraction from snow," U.S. Patent Application US007 654 260B2, Feb. 2, 2010.

[13] Honeywell, "Honeywell Zephyr Analog Airflow Sensors: HAF Series-High Accuracy," https://docs.rs-online.com/c282/0900766b813e6b3e.pdf, [Online; visited on 02 April, 2021].

[14] Bosch Sensortec, "BME680, Low power gas, pressure, temperature & humidity sensor (Rev. 1.6)," https://www.bosch-sensortec.com/products/environmental-sensors/gas-sensors/bme680/#documents, January 2021, [Online; visited on 06. February, 2021].

Armin Stockhammer is an Engineer at MED-EL Medical Electronics, Manufacturing Engineering, External Devices. He acts as an interface between R&D and Manufacturing by supporting the product developments and their life-cycles.

Entwicklung einer Ausbildungsplattform für serielle Industrieroboterkinematiken

Christoff Sulzenbacher B.Sc., Benjamin Massow B.Sc. M.Sc. (Betreuer)

***Kurzfassung*—Die Robotik ist in der heutigen technischen Welt unabdingbar. Fehlende Fachkräfte zwingen die Industrie zur immer weiteren Automatisierung, wodurch die Nachfrage nach gut ausgebildeten Technikern, die mit der Materie Robotik und Automatisierung vertraut sind, steigt. Oft sind jedoch junge Techniker, die nach ihrer akademischen Ausbildung in das Berufsleben einsteigen, mit der Praktik wenig vertraut. Dies liegt oft am Fokus der theoretischen Ausbildung. Um dem entgegenzuwirken ist den auszubildenden Personen einen praktisch-orientierter Zugang zur Robotik zu verschaffen. Hierzu ist es notwendig, eine Ausbildungsplattform zu schaffen, die es erlaubt, tiefer in die Kinematik und Dynamik einzugreifen, als dies bei kommerziellen Robotern der Fall ist. Es wird anhand der im Unterricht vermittelten Lerninhalte ein Unterrichtskonzept erstellt, welches den Fokus auf die Integration einer Roboteranlage legt. Diese Ausbildungsplattform wird als Selective Compliance Articulated Robot Arm (SCARA) ausgelegt. Die Ansteuerung wird über eine speicherprogrammierbare Steuerung (SPS) der Firma Beckhoff implementiert. Mithilfe einer Bedienoberfläche lässt sich der Roboter einfach ansteuern. Neben der reellen Anlage wird eine virtuelle Ausbilungsplattform erstellt, die es den Kursteilnehmern ermöglicht, eigens erstellte Kinematiken und Trajektorien zu testen, ohne dabei eine Hardware beschädigen zu können. Ein hypothetischer Durchlauf der Unterrichtsmethodik, gekoppelt mit der Ausbildungsplattform, führt zu dem Schluss, dass ein solch gekoppeltes System der Robotikausbildung erheblichen Mehrwert bietet und zur besseren Vermittlung der Lerninhalte führt.**

***Schlagwörter*—Roboter, Ausbildung, Steuerung**

C. Sulzenbacher arbeitet bei INNIO Jenbacher GmbH & Co. OG, Jenbach, Österreich, e-mail: ch.sulzenbacher@mci4me.at.

B. Massow arbeitet am Zweig Mechatronik am MCI, Innsbruck, Österreich.

Manuskript eingereicht am 29. Juli 2022

I. Einleitung

DIE Robotik weist in der Industrie eine immer größere Wichtigkeit auf. Der Fachkräftemangel, der in den letzten Jahren verstärkt zu spüren ist, erfordert gezieltes Gegensteuern, in Form von vermehrter Automatisierung [1]. Der Berufseinstieg fällt vielen angehenden Robotik- und Automatisierungspezialisten oft schwer, da sich die theoretisch vermittelten Inhalte der Ausbildung schwer mit dem praktisch relevanten Wissen verknüpfen lassen. Es ist somit im Interesse der ausbildenden Institute deren Abgängern einen möglichst engen Bezug zwischen Theorie Praxis zu vermitteln. Für den Fall der Robotik ist es notwendig, eine Ausbildungsplattform zu schaffen, die es erlaubt, tiefer in die Kinematik und Dynamik einzugreifen, als dies bei einem kommerziellen Roboter der Fall ist. Dieser Artikel stellt sich somit der Frage, ob es möglich ist, solch eine Ausbildungsplattform für serielle Kinematiken zu erstellen, die, zusammen mit einem ausgearbeiteten Unterrichtskonzept, die gewünschten Lerninhalte erfolgreich vermitteln kann. Dieser Artikel ist in Koordination mit den Arbeiten *Entwicklung einer Ausbildungsplattform für serielle Industrieroboterkinematiken* und *Entwicklung einer Ausbildungsplattform zum modellbasierten Entwicklungsprozess von Automatisierungslösungen am Beispiel von Industrieroboterkinematiken* entstanden [2], [3].

II. Theoretische Grundlagen

A. Selective Compliance Articulated Robot Arm (SCARA)

Im Gegensatz zu gewöhnlichen 6-Achs-Knickarmen besitzt der SCARA nur vier Freiheitsgrade. Drei davon werden über Rotationsachsen definiert, eine über eine Linearachse. Durch die parallele Ausrichtung aller Achsen, ergibt sich ein zylindrischer Arbeitsbereich. Der Nachteil des verringerten Arbeitsraums wird durch eine höhere Positioniergenauigkeit in x und y kompensiert, wodurch sehr hohe Geschwindigkeiten und Beschleunigungen erreicht werden können [4].

B. Kinematisches Modell

1) Relative Posen: Um einen Gegenstand im Raum richtig zu beschreiben, sind sechs Parameter notwendig, drei für die Position des Gegenstandes, sowie drei für dessen Rotation. Wird dieser Gegenstand mit diesen sechs Parametern beschrieben, handelt es sich um eine Pose. Da sich diese Pose auf ein bestimmtes Koordinatensystem bezieht, handelt es sich hierbei um eine relative Pose. In der Robotik wird, um den Bezug von einem Koordinatensystem $\{A\}$ auf ein anderes $\{B\}$ darzustellen, der griechische Buchstabe ξ verwendet. Wenn das Koordinatensystem $\{B\}$ in Bezug zu $\{A\}$ angegeben wird, so wird dies ausgedrückt als ${}^{A}\xi_{B}$. Um eine Pose in mathematischer Form repräsentieren zu können, wird diese in die Komponenten Rotation und Translation zerlegt und einzeln behandelt. Eine Rotation von einem Koordinatensystem $\{A\}$ in ein anderes $\{B\}$ kann mithilfe einer Orthonormalmatrix ${}^{A}\mathbb{R}_{B}$ realisiert werden. Eine Translation von einem Koordinatensystem in ein anderes kann mithilfe eines Vektors beschrieben werden, solange diese Koordinatensysteme parallel zueinander stehen. Dies ist immer der Fall, wenn das Ursprungskoordinatensystem zuerst in die Orientierung des zweiten Koordinatensystems rotiert wird. Damit die Rotation mit der Translation kombiniert werden kann, wird die orthonormale Rotationsmatrix um einen translativen Term erweitert. Es ergibt sich

$$ {}^{A}\xi_{B} \sim^{A} \mathbb{T}_{B} = \begin{bmatrix} {}^{A}\mathbb{R}_{B} & {}^{A}t_{B} \\ \mathbb{O}_{1x3} & 1 \end{bmatrix} , \tag{1} $$

wobei $\mathbb{O}$ dem Vektor $[0\ 0\ 0]$ entspricht.

C. Vorwärtskinematik

1) Denavit-Hartenberg Parameter: Um zu wissen, wo sich der Endeffektor bei einer gewissen Achsposition der Motoren befindet, ist es notwendig, die Vorwärtskinematik des Robotermodells zu lösen. Für mehrdimensionale serielle Roboter im dreidimensionalen Raum wird dies über die DH-Parameter gelöst [5].

Der Vorteil der DH-Notation ist, dass der Zusammenhang von zwei auf den Roboterachsen liegenden Koordinatensystemen mit bloß vier Parametern bestimmt werden kann. Um die DH-Parameter eines Roboters zu definieren, wird dieser in n Achsen und $n+1$ Gelenke unterteilt, wobei das Gelenk 0 gleich der Basis des Roboters entspricht. Anschließend wird dem distalen Ende jedes Gelenks ein Koordinatensystem angefügt, dessen z-Achse bei rotatorischen Achsen immer entlang der Rotationsachse verläuft und bei Linearachsen immer entlang der Translationsrichtung. Die Pose $\{i\}$ in Bezug auf $\{i-1\}$ wird in der DH-Notation beschreiben durch

$$ {}^{i-1}\xi_{i} \sim \mathbb{A} = \mathcal{R}_{z}(\theta_{i}) \oplus \mathcal{T}_{z}(d_{i}) \oplus \mathcal{T}_{x}(a_{i}) \oplus \mathcal{R}_{x}(\alpha_{i}) , \tag{2} $$

wobei $\mathcal{R}$ und $\mathcal{T}$ jeweils elementarer Rotationen und Translationen entsprechen [6]. Abhängig vom Achsentyp ist entweder der Wert θ_i (rotatorisch) oder der Wert d_i (linear) gleich dem Achswert. a_i,d_i,θ_i und α_i sind dabei konstante Werte, abhängig von der Pose i. Diese Beschreibung einer relativen Pose mit nur vier Parametern ist möglich, da für diese Achsen zwei Bedingungen eingehalten werden müssen:

- Die x-Achse des Koordinatensystems i kreuzt die z-Achse des Koordinatensystems $i-1$.
- Die x-Achse des Koordinatensystems i steht senkrecht auf die z-Achse des Koordinatensystems $i-1$.

Die DH-Parameter eines zweidimensionalen Roboters mit zwei Achsen sind in Tabelle I zu sehen.

Tabelle I
DH-PARAMETER EINES ZWEIACHSROBOTERS IN DER EBENE.

θ_i	d_i	a_i	α_i
θ_1	0	a_1	0
θ_2	0	a_2	0

2) Analytische Berechnung der Vorwärtskinematik: Betrachtet man einen SCARA der sich nur in einer Ebene bewegt - die Linearachse ist als Schalter anzusehen - ist es möglich, die Vorwärtskinematik auf algebraische Art herzuleiten. Eine arbiträre Pose mit den bekannten Winkeln und Längen des Roboters ist in Abbildung 1 zu sehen.

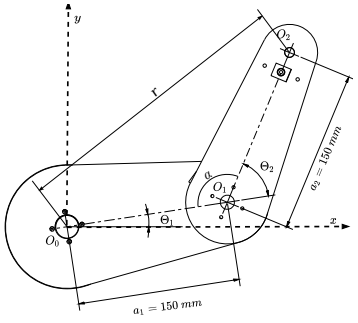


Abbildung 1. Eine arbiträre Pose des SCARA im zweidimensionalen Raum. Die Basis O_0 entspricht gleich der ersten Achse des Roboters und sitzt im Ursprung des Koordinatensystems. Das Koordinatensystem der zweiten Achse liegt in O_1, das des Endeffektors in O_2.

Die Position des Endeffektors lässt sich geometrisch aus der Abbildung 1 beschreiben durch

$$\begin{aligned} x =& a_1 \cos(\theta_1) + a_2 \cos(\theta_1 + \theta_2) \\ y =& a_1 \sin(\theta_1) + a_2 \sin(\theta_1 + \theta_2) \end{aligned} \quad . \tag{3}$$

a_1 entspricht dabei der Länge des ersten Armes, a_2 die des zweiten Armes. Bei θ_1 und θ_2 handelt es sich jeweils um die beiden Achswinkel.

3) Analytische Lösung der inversen Kinematik: Anhand der Abbildung 1 lässt sich erkennen, dass mithilfe des Winkels α und der Länge r, θ_2 berechnet werden kann als

$$\theta_2 = \arctan_2(\pm \frac{\sqrt{1 - \cos(\theta_2)}}{\cos(\theta_2)}) \ . \tag{4}$$

Um den zweiten Winkel zu bestimmen, wird das Quadrat der Gleichungen in 3 genommen und zusammengezählt, womit sich für x und y in Matrixform ausgedrückt

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} a_1 + a_2 \cos(\theta_2) & -a_2 \sin(\theta_2) \\ a_2 \sin(\theta_2) & a_1 + a_2 \cos(\theta_2) \end{bmatrix} \begin{bmatrix} \cos(\theta_1) \\ \sin(\theta_1) \end{bmatrix} \tag{5}$$

ergibt [7]. Da nach $\cos(\theta_1)$ und $\sin(\theta_1)$ aufgelöst werden muss, kann die Cramersche Regel angewandt werden.

θ_1 kann somit berechnet werden über

$$\theta_1 = \arctan_2 \left(\frac{\sin(\theta_1)}{\cos(\theta_1)} \right) \tag{6}$$

4) Singularität: Eine Singularität in der Robotik ist dadurch gekennzeichnet, dass ein definierter Punkt im kartesischen Raum sich nicht eindeutig einer Achsposition im Gelenksraum zuordnen lässt. Am Beispiel eines SCARA ist dies der Fall, wenn die zweite Achse um π verdreht ist, beziehungsweise der Endeffektor sich im Ursprung befindet. Diese Art der Singularität ist eine innere Singularität. Falls ein Punkt nicht erreicht werden kann, da dieser außerhalb des Arbeitsbereichs des Roboters liegt, wird dies als äußere Singularität bezeichnet.

D. Trajektorienplanung

Die Trajektorienplanung ist dafür zuständig, einen Bezug zwischen Zeit t und Ort p herzustellen. Da die Trajektorienplanung vor Allem für die Motoren relevant ist, wird auf die eindimensionale Trajektorienplanung eingegangen. Des Weiteren werden nur Punkt-zu-Punkt Trajektorien betrachtet, da ausschließlich solche in der Ausbildungsplattform implementiert werden. Es gibt eine Vielzahl an Methoden der Trajektorienimplementierung, die, abhängig

von der durchzuführenden Aufgaben, weniger bis sehr komplex sein können. Für die Ausbildungsplattform wird diese jedoch nur mit der in der Industrie am häufigsten verwendeten Methode, der Trajektorienplanung mit trapezförmigen Geschwindigkeitsprofil, umgesetzt. Für weitere Bewegungsprofile wird auf weiterführende Literatur verwiesen [8], [9].

III. Konzept zur Einbindung in den Unterricht

Ziel der Entwicklung einer Ausbildungsplattform für serielle Kinematiken ist es, den Kursteilnehmern die Grundlagen der Robotik schrittweise näher zu bringen. Das Lernziel am Ende des Kurses soll sein, dass ein kinematisches Modell, eines zweidimensionalen seriellen Roboters, inklusive inverser und Vorwärtskinematik, erstellt werden kann. Zusätzlich sollen die Kursteilnehmer am Ende in der Lage sein, eine Trajektorie zwischen zwei Punkten zu erzeugen, sowohl im Achsraum, als auch im kartesischen Raum. Ebenfalls soll ein Verständnis über die Problematiken von Singularitäten vorhanden sein.

A. Einführung in die Robotik

Zu Beginn des Unterrichts ist es von großer Bedeutung, eine gemeinsame Basis für die Kursteilnehmer zu schaffen, damit ein annähernd gleicher Wissensstand vorliegt. Dazu dient eine kurze Einführung von unterschiedlichen Roboterarten und deren Einsatzgebiete in der Industrie. Dabei werden Begriffe wie *serielle Kinematik*, *parallele Kinematik*, *Freiheitsgrade* und *Arbeitsbereich* erläutert.

B. Erklärung von relativen Posen

Im Anschluss wird die Bedeutung von relativen Posen näher erläutert. Neben der Idee hinter den Posen wird die Berechnung dieser relativen Posen erklärt. Die orthonormale Rotationsmatrix wird erklärt, genauso wie die homogene Transformationsmatrix. Damit bei der Berechnung einer Endeffektorposition mithilfe von Posen der Aufwand nicht zu groß wird, wird dies nur mit einem Zweiachsroboter in der Ebene besprochen.

C. Erstellen eines kinematischen Modells

Im Anschluss wird die Vorwärtskinematik mithilfe der DH-Parameter erklärt. Dabei sollen am Ende die DH-Parameter eines Zweiachsroboters in der Ebene selbstständig gefunden werden. Aufbauend auf die Vorwärtskinematik des Zweiachsroboters wird eine trigonometrische Lösungsmethode der inversen Kinematik besprochen. Dabei soll aufgezeigt werden, dass es in der Lösung mathematisch nicht lösbare Positionen gibt. Diese Punkte sollen ermittelt und graphisch dargestellt werden. In diesem Zusammenhang soll der Begriff der *Singularität* erklärt werden. Um die hergeleitete Kinematik im Robotermodell testen zu können, wird sowohl die Vorwärtskinematik als auch die inverse Kinematik als *Matlab*-Skript verfasst. Diese können als verwendete Kinematik in der virtuellen Ausbildungsplattform eingepflegt und getestet werden.

1) Einbindung der Kinematik in eine virtuelle Ausbildungsplattform: Um zu überprüfen, ob die - in den zuvor durchgeführten Übungen - erzeugten *Matlab*-Funktionen wie gewünscht funktionieren, können sie in einem virtuellen Robotermodell eines Zweiachsroboters eingepflegt werden. Das in *Simulink* erzeugte Robotermodell fährt eine vordefinierte Trajektorie ab und verwendet dabei die von den Kursteilnehmern erstellen kinematischen Funktionen. Wird dem virtuellen Robotermodell keine Funktion eingepflegt, wird auf intern hinterlegte, korrekte kinematische Modelle zurückgegriffen.

D. Trajektorienplanung

Abhängig von der Aufnahmefähigkeit und Lerngeschwindigkeit der Kursteilnehmer kann zusätzlich zur Kinematik noch die Trajektorienplanung behandelt werden. Diese wird nach erfolgreicher Implementierung der Kinematik in der virtuellen, sowie der reellen Ausbildungsplattform behandelt. Hier werden Begriffe wie *Pfad*, *Trajektorie*, *kartesischer Raum* und *Achsraum* erläutert. Es wird ein Einblick über die Erzeugung von Bewegungsprofilen, die Abhängigkeiten zwischen Position, Geschwindigkeit und Be-

schleunigung, sowie über die mathematische Herleitung einer Trajektorie mit trapezförmigen Geschwindigkeitsprofil geboten. Anschließend werden die Unterschiede zwischen einer Linearbewegung und einer Achsbewegung erläutert. Abschließend wird die Synchronisation von mehreren Achsen beschrieben.
Es wird den Teilnehmern selbst überlassen, die theoretischen Konzepte in ein funktionsfähiges *Matlab*-Script zu transferieren. Dabei sollen die bereits integrierten Funktionen zur Trajektorienplanung aus der *Robotic System*-Toolbox verwendet werden, um den Aufwand der Programmierung so gering wie möglich zu halten. Zum Testen der erzeugten Trajektorien, sowohl für Linear- als auch Achsbewegungen, können diese in die virtuelle Ausbildungsplattform eingebunden werden. Des Weiteren soll die Möglichkeit gegeben sein, eine bereits extern generierte Trajektorie mit beliebig vielen Zwischenpunkten als Datei der virtuellen Ausbildungsplattform zu übergeben, die diese ausführt.

E. Einbinden eines reellen Robotermodells in Form einer Ausbildungsplattform

Das abschließende Ziel ist es, die an der virtuellen Ausbildungsplattform getesteten Funktionen an einer reellen Hardware zu testen. Hierfür ist es notwendig, dass die virtuelle Ausbildungsplattform ident ist mit einer reellen Ausbildungsplattform. Da für die Herleitung der Kinematik speziell auf den Zweiachsroboter eingegangen wird, ist die Realisierung einer Hardware mit denselben Eigenschaften erwünscht. Es wird somit für den reellen Roboter ein SCARA entwickelt, der aus zwei Rotationsachsen besteht, sowie einer Linearachse, die nur aus- und eingefahren werden kann. Dadurch muss die Trajektorienplanung nur in der Ebene betrachtet werden.

IV. Umsetzung

Eine Ausbildungsplattform für serielle Roboterkinematiken könnte auf verschiedene Arten realisiert werden. Da als Vorgabe gegeben war, dass die fertiggestellte Anlage auf einem bereits vorhandenen Tisch platz haben soll, wird die Konstruktion auf die Dimensionen dieses Tisches ausgelegt. Da sich die Ausbildungsplattform an Personen richtet, die kein Vorwissen bezüglich Roboterprogrammierung vorweisen, ist es erwünscht, die Kinematik des Systems so einfach wie möglich zu halten, weshalb ein SCARA ausgewählt wird.
Das Gehäuse des Roboters wird aus Aluminiumprofilen konstruiert, das auf die Größe der Tischplatte ausgelegt ist. Es werden im obersten Viertel des Würfels Querstreben eingefügt, die als Halterung der Elektronik dienen. Des Weiteren werden vertikal vier weitere Aluminiumprofile eingebaut, auf denen im Anschluss der Roboter montiert wird.
Die Arme des Roboters sind mit jeweils $150\,\mathrm{mm}$ dimensioniert. Der Radius des End-Effektors liegt bei $600\,\mathrm{mm}$, wodurch der Arbeitsraum die Dimension der Einhausung nicht überschreitet. Die Arme sind aus Polymethylmethacrylat (Plexiglas) gefertigt und sind an die Form der verwendeten Motoren angelehnt. Die Dimensionen der Arme, als auch eine Gesamtübersicht der Anlage sind in Abbildung 2 zu sehen.

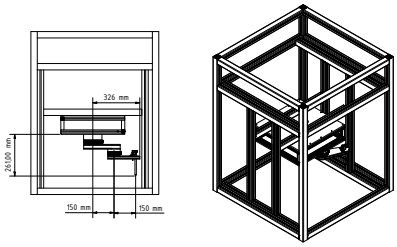


Abbildung 2. Abmaße des Roboters, sowie die der Arme. Mit einem Radius von $326\,\mathrm{mm}$ ist der Durchmesser geringer als der des Rahmens, somit ist keine Kollision mit der Umgebung möglich.

A. Komponentenauswahl

1) Motoren: Für die Achsen werden zwei Gimbal-Motoren der Firma *T-Motor* ausgewählt, für die Basis den Typ *G-100*, für die zweite Achse den Typ *G-60*.

Diese weisen ein hohes Drehmoment auf und besitzen eine Hohlwelle, worin Schleifringe montiert werden können, die zur Durchführung von Versorgung- und Steuerleitungen dienen.

2) Sensoren: Für die Positionsregelung werden jedem Motor jeweils ein Absolut- und ein Inkrementalgeber der Firma *RLS* angefügt. Beim Inkrementalgeber handelt es sich um das Modell *RLC2ICA13BC00A00*, mit dem dazu passenden Encoderring *MR047B*. Der Ring weist einen Außendurchmesser von $(47.5 \pm 0.1)\,\mathrm{mm}$ auf, mit einer Pollänge von $2\,\mathrm{mm}$. Es ergeben sich dadurch für den Ring eine Gesamtpolzahl von 76 [10].

Beim Absolutgeber handelt es sich um das Modell *MB049SCB19BENT00*, mit dem passenden Magnetring *MRA049BC025DSE00*. Die Dimension des Absolutgebers ist auf die Größe des Relativgeberrings abgestimmt, um leichter miteinander verbaut werden zu können. Für die Kommunikation wird das SSI-Übertragungsprotokoll verwendet, welches von *Beckhoff* empfohlen wird.

3) Steuerung: Zur Steuerung der Anlage wird ein Industrie-PC (IPC) der Firma *Beckhoff* verbaut. Zur Bedienung der *Windows*-Betriebsoberfläche ist an der Seite der Anlage das *Beckhoff*-Bedienpanel *CP6919* verbaut. Dieses Single-Touch-Display kann zum Einen für das Bedienen der Human-Machine-Interface (HMI) verwendet werden, zum Anderen um programmtechnische Änderungen vorzunehmen. Für größere Änderungen kann über ein Virtual Private Network (VPN) auf den IPC zugegriffen werden. Ein Bedienpanel des Typs *CP2924* der Firma *Beckhoff*, welches als Eingabegerät für den Roboter dient, kann ebenfalls auf dem IPC angeschlossen werden. Die Feldbusklemmen werden mithilfe des Buskopplers *EK1100* an den IPC angeschlossen. An diese werden die Feldbusklemmen EL4711, EL5001, sowie EL2008 angeschlossen. Bei der EL4711 handelt es sich um die Motorklemme für die Ansteuerung der beiden Motoren. Die SSI-Geber Klemme EL5001 ist für die Auswertung der Absolutgeber zuständig, während es sich bei der EL2008 um eine Digitale Ein-Ausgangsklemme handelt, die zur Ansteuerung des End-Effektors verwendet wird.

B. Virtuelle Ausbildungsplattform

Da eine Ausbildungsplattform dafür gedacht ist, neu gelerntes auszuprobieren, ist es wichtig, dass diese nicht bei einem Programmfehler des Nutzers zunichte geht. Aus diesem Grund erhält die mechanische Ausbildungsplattform ein virtuelles Ebenbild, welches das gleiche Verhalten wie die reelle Anlage aufweist. Es wird in diesem Artikel ausschließlich auf die Implementierung der seriellen Ausbildungsplattform, mit Fokus auf die Kinematik der Anlage, eingegangen. Für weiterführende Informationen über den Datentransfer und Kommunikation zwischen HMI, *Beckhoff*, *Simulink* und weitere, ist die Arbeit *Entwicklung einer Ausbildungsplattform zum modellbasierten Entwicklungsprozess von Automatisierungslösungen am Beispiel von Industrieroboterkinematiken* heranzuziehen [2].

V. Used-Case

Der Used-Case soll aufzeigen, ob das erstellte Unterrichtskonzept, mithilfe der Lernplattform, die gewünschten Lerninhalte vermitteln kann. Dafür wird der Ablauf des Unterrichts simuliert.

A. Umsetzbarkeit der zu vermittelnden Lerninhalte

1) Einführung in die Robotik: Die Kursteilnehmer haben im Unterricht die Grundbegriffe der Robotik gelernt und sind in der Lage die ihnen gegebenen Übungen, zur Einführung in die Robotik, zu lösen. Den Kursteilnehmern ist das Lösen der gestellten Aufgaben mithilfe der im Unterricht vermittelten Inhalte gelungen und es kann das nächste Themengebiet angegangen werden.

2) Erklärung von relativen Posen: Den Kursteilnehmern, die bisher wenig Erfahrung mit Koordinatensystemen hatten, können mithilfe von ausgehändigten Koordinatensystemen dem Unterricht leichter folgen und es kann mit dem Themengebiet der relativen Posen schneller begonnen werden. Im Unterricht sind die relativen Posen behandelt worden

wodurch den Kursteilnehmern die Rechenregeln und der gerichtete Graph ein Begriff sind. Zum Festigen des neu Erlernten wird die dazu vorgesehene Übung gelöst. Die Kursteilnehmer sich den Umgang mit relativen Posen angeeignet. Es wird nun die Rotationsmatrix erklärt und es sind die Übungen zu Rotationsmatrizen zu lösen.
Das Verständnis für Rotationsmatrizen ist von den Kursteilnehmern gefestigt worden und es kann somit die homogene Transformationsmatrix besprochen werden. Die Translation von Koordinatensystemen ist verständlich erklärt worden und es kann die dazugehörende Übung gelöst werden.

3) Erstellen eines kinematischen Modells: Der Unterricht vermittelt den Teilnehmern, wofür die Vorwärtskinematik bei Robotern benötigt wird. Des Weiteren wird erläutert, wie sie berechnet werden kann, sowohl mithilfe von DH-Parametern, als auch auf analytische Weise. Zur Wissensüberprüfung sind die Übungen zur Vorwärtskinematik zu lösen. Sobald dies den Kursteilnehmern gelungen ist, kann der Unterricht sich weiter mit der inversen Kinematik beschäftigen. Den Teilnehmern wird erklärt, weshalb die inverse Kinematik benötigt wird und wie sie für einfache Aufbauten zu berechnen ist. Zum Festigen des Verständnisses über inverse Kinematiken soll dazu eine Übung gelöst werden.

Durch die Erstellung der Funktion für die inverse Kinematik der Ausbildungsplattform, ist es den Kursteilnehmern nun möglich, die selbst geschriebene Kinematik an der virtuellen Ausbildungsplattform auszuprobieren. Dafür wird anhand der vorgegebenen Anleitung die virtuelle Ausbildungsplattform in Betrieb genommen und ein Gefühl für die Programmoberfläche gegeben. Bei der Inbetriebnahme haben die Teilnehmer sich mit der Oberfläche vertraut gemacht und sind in der Lage, den Roboter mit der vorgegebenen Steuerung zu bedienen. Anschließend wird mithilfe der Anleitung die eigene Vorwärtskinematik, als auch die inverse Kinematik eingepflegt. Den Teilnehmern ist es gelungen, die eigens berechneten Kinematiken in der virtuellen Ausbildungsplattform einzupflegen, um diese zu evaluieren.

4) Trajektorienplanung: Damit der Roboter sich zwischen vordefinierten Punkten bewegen kann, ist es notwendig, eine Trajektorie zwischen den Punkten zu erzeugen. Die dafür erstellten Übungen fordern die Erstellung eines *Matlab*-Skripts, die eine solche Trajektorie erstellt - einmal durchgeführt durch eine Achsbewegung, und einmal anhand einer Linearbewegung. Hierbei wird kein dezidierter Lösungsvorschlag angegeben, da je nach Aufnahmefähigkeit der Kursteilnehmer die Trajektorie entweder selbst berechnet wird, oder lediglich mit der in *Matlab* vorhandenen Funktion eingepflegt wird. Dieses Skript speichert die einzelnen Trajektorienpunkte in eine `.csv`-Datei.

Den Kursteilnehmern ist es gelungen, ein `.csv`-Dokument zu erstellen, in dem die einzeln aufgeschlüsselten Punkte der gewünschten Trajektorie zu sehen sind. Im Anschluss ist die virtuelle Ausbildungsplattform zu starten und über die HMI die Trajektorie als `.csv`-Datei hochzuladen. Durch das anschließende Drücken der Play-Taste, lässt sich der Verlauf des Roboters beobachten.
Es wurde möglich, den Unterschied zwischen linearer Trajektorien und Achstrajektorien in der virtuellen Ausbildungsplattform erkennbar zu machen und somit das Verständnis über Trajektorien zu verbessern. Um nun die Trajektorie am reellen Roboter testen zu können, wird die Ausbildungsplattform in Betrieb genommen. Durch die Schritt-für-Schritt Anleitung ist die Inbetriebnahme der Ausbildungsplattform für die Kursteilnehmer ohne Vorwissen über die Anlage möglich.
Es ist den Kursteilnehmern gelungen den Roboter in Betrieb zu nehmen und einfache Jog-Bewegungen über die HMI vorzunehmen. Als Endziel des Kurses ist nun die eigene Trajektorie auf der reellen Ausbildungsplattform abzufahren.
Den Teilnehmern ist es gelungen, die Trajektorie abfahren zu lassen, und somit den Kurs erfolgreich abzuschließen.

Der fiktive Durchlauf des Unterrichtkonzepts macht ersichtlich, dass sich der Einsatz der entwickelten Ausbildungsplattform für serielle Kinemati-

ken für das Vermitteln der Grundlagen der Robotik eignet.

VI. Conclusio

Ziel der Arbeit *Ausbildungsplattform für serielle Kinematiken* ist es, mit einem Unterrichtkonzept einen möglichst intuitiven Einstieg in die Materie der Robotik zu geben. Dabei soll die Theorie mit der Praxis Hand in Hand gehen. Um dies zu ermöglichen, wurde eine Roboterzelle konstruiert, die einen Roboter beinhaltet. Dieser ist ähnlich eines SCARA aufgebaut, um die Kinematik des Roboters so einfach wie möglich zu halten. Mittels zweier Touchbildschirme kann der Bediener einerseits den Roboter bedienen, andererseits kann der Roboter mithilfe des am Endeffektor montierten Stifts mit der Oberfläche interagieren.

Zusätzlich zur reellen Ausbildungsplattform wird eine virtuelle Ausbildungsplattform erstellt, die sich ident zur reellen Anlage verhält. Die in *Matlab Simulink* implementierte Zelle lässt sich mit derselben HMI ansteuern wie die reelle Anlage. Dies ermöglicht einen fließenden Übergang zwischen Simulation und Realität. Die virtuelle Ausbildungsplattform erlaubt es, selbst erstellte *Matlab*-Funktionen zu integrieren, die die Vorwärts- und inverse Kinematik der Anlage beschreiben. Wird mit der eigenen Kinematik eine Trajektorie berechnet und die einzelnen Achswerte zu fixen Zeitschritten in einer `.xml`-Datei gespeichert, so kann diese nicht nur am virtuellen, sondern auch am reellen Roboter abgefahren werden.

Der Used-Case in Abschnitt V zeigt deutlich, dass das Konzept für den Unterricht sich mithilfe der Ausbildungsplattform umsetzen lässt. Somit lässt sich die Frage, ob es möglich ist, solch eine Ausbildungsplattform für serielle Kinematiken zu erstellen, die zusammen mit einem ausgearbeiteten Unterrichtskonzept die gewünschten Lerninhalte erfolgreich vermitteln kann, mit ja beantworten. Das Konzept, welches für den Unterricht in dieser Arbeit ausgearbeitet wurde, kann abhängig von den der Lehrperson gesetzten Schwerpunkten etwas abgeändert werden, die Integrierbarkeit der Ausbildungsplattform bleibt dennoch bestehen.

Danksagung

Der Autor möchte sich bei seinem Betreuer Benjamin Massow bedanken, der stets für Fragen zur Seite stand und mit konstruktivem Feedback die Arbeit auf die Zielgerade brachte.

Literatur

[1] M. Ciopek, "Automatisierungsbranche verzeichnet starken auftragszuwachs," 11 2021.

[2] J. Muigg, "Entwicklung einer ausbildungsplattform zum modellbasierten entwicklungsprozess von automatisierungslösungen am beispiel von industrieroboterkinematiken," 2021.

[3] S. Kollegger, "Entwicklung einer ausbildungsplattform für parallele industrieroboterkinematiken," 2021.

[4] E. Uhlmann and J. Krüger, *Industrieroboter*, 11 2020, pp. 1205–1218.

[5] P. Corke, *Robotics,Vision and Control*, 2013, pp. 197–200.

[6] B. Massow, *Applied Robotics - Condensed Theory*.

[7] C. Dede, *Inverse Kinematics Analysis of RR Manipulator*, 2020, Management Center Innsbruck.

[8] L. Biagiotti and C. Melchiorri, *Trajectory Planning for Automatic Machines and Robots*, 2008.

[9] M. Ceccarelli, *Fundamentals of Mechanics of Robotic Manipulation*, ser. Intelligent Systems, Control and Automation: Science and Engineering. Springer Netherlands, 2013.

[10] R. meilna tehnika d.o.o., *Radial incremental magnetic rings*, 07 2020.

Christoff Sulzenbacher studiert Mechatronik & Smart Technologies am MCI Innsbruck/Österreich. Nebenbei arbeitet er als Robotikexperte bei INNIO Jenbacher GmbH & Co. OG.

Modellierung der Wärme- und Massenübertragung eines zur additiven Fertigung von Keramikbauteilen verwendeten Trocknungsprozesses mittels numerischer Strömungssimulation

Oliver Ungerer, Clemens Pörnbacher (Betreuer) und Martin Pillei (Begutachtender)

***Kurzfassung*—Der Einfluss der Trocknung der frisch aufgetragenen Materialschicht ist von großer Relevanz im Bezug auf Produktivität und Materialqualität. Dies bezieht sich vor allem auf die additive Fertigung mittels wasserbasierten Schlickern.**

Die Herstellung von Grünlingen aus Aluminiumoxid weist dabei eine hohe Anfälligkeit für Risse und Fehlstellen auf. Die Entfernung eines hohen Anteils von Wasser in kurzer Zeit bei gleichzeitig geringen Temperatur- und Feuchtegradienten im Bauteil stellt die Zielsetzung der Prozessentwicklung dar.

Der Trocknungsprozess wird simulativ in OpenFOAM umgesetzt. Es kommt eine modifizierte Form des Solvers chtMultiRegionFoam in Kombination mit einer selbstentwickelten Randbedingung zur Anwendung. Zur Validierung des Modells werden Simulationsergebnisse mit Messungen abgeglichen. Der Abgleich wird in Bezug auf Masseabnahme und Oberflächentemperatur durchgeführt. Die erhaltenen Ergebnisse zeigen, dass die Anwendung von CFD zur Abbildung von Trocknungsprozessen geeignet ist. Sie gibt detaillierte Einsicht in den Ablauf und das Zusammenspiel gekoppelter physikalischer Prozesse und kann als leistungsstarkes Werkzeug zur Prozessoptimierung eingesetzt werden.

O. Ungerer studiert am Studiengang Mechatronik, MCI, Innsbruck, Österreich, e-mail: oungerer.@mci4me.at

C. Pörnbacher arbeitet bei der Durst Group AG in der technischen Entwicklungsabteilung, Brixen, e-mail: Clemens.Poernbacher@durst-group.com

M. Pillei hat eine Professur am MCI, Innsbruck, Österreich, e-mail: Martin.Pillei@mci.edu

Manuskript eingereicht am 29. Juli 2022; revidiert am 29. Juli 2022.

***Schlagwörter*—Additive Fertigung, Technische Keramik, Trocknung, OpenFOAM**

SYMBOLVERZEICHNIS

$\vec{u}$	Geschwindigkeit in $\mathrm{m\,s^{-1}}$
T	Temperatur in K
p	Druck in Pa
Y_{H_2O}	Massenanteil Wasserdampf in Luft $\mathrm{kg\,kg^{-1}}$
Y_W	Massenanteil Wasser in Aluminiumoxid $\mathrm{kg\,kg^{-1}}$
ρ	Dichte in $\mathrm{kg\,m^{-3}}$
h	Enthalpie in $\mathrm{J\,kg^{-1}}$
h_{lat}	Verdampfungsenthalpie in $\mathrm{J\,kg^{-1}}$
$\dot{Q}$	Wärmestrom in W
$\dot{q}$	flächenspezifischer Wärmestrom in $\mathrm{W\,m^{-2}}$
g	Gravitationsbeschleunigung in $\mathrm{m\,s^{-2}}$
α	Diffusivität in $\mathrm{m^2\,s^{-1}}$
μ	kinematische Viskosität in $\mathrm{m^2\,s^{-1}}$
D_Y	Diffusivität Wasser in Aluminiumoxid in $\mathrm{m^2\,s^{-1}}$
λ	Wärmeleitfähigkeit in $\mathrm{W\,m^{-1}\,K^{-1}}$
c_p	spezifische Wärmekapazität in $\mathrm{J\,kg^{-1}\,K^{-1}}$
$\dot{m}$	Massenstrom in $\mathrm{kg\,m^{-2}\,s^{-1}}$
m	Masse in kg
ξ	Massenanteil Aluminiumoxid in $\%$
a_W	Wasseraktivität in $-$
RH	Relative Luftfeuchte in $-$
$\vec{F}$	Kraft in N
k	turbulente kinetische Energie in $\mathrm{m^2\,s^{-2}}$
ω	spezifische turbulente Disspationsrate in $\mathrm{s^{-1}}$

I. Einleitung

Die additive Fertigung von Bauteilen aus technischer Keramik ist für kleine, dünnwandige Bauteile durch Verwendung litografiebasierter Verfahren etabliert [1], [2]. Die Herstellung großer, dickwandiger Bauteile ist aufgrund des notwendigen Entbinderungsschritts nicht prozesssicher möglich. Alternative Verfahren mit einem hohem Wasser- und sehr geringem Organikanteil des Schlickers schaffen Abhilfe [3].
Das Entfernen der Feuchtigkeit aus dem schichtweise aufgebrachten Schlicker durch Trocknung stellt einen kritischen Teil im Fertigungsprozess dar. Inhomogene Trocknung führt zur Bildung von Rissen und Verformung der Bauteile und mindert deren Qualität. Die messtechnische Erfassung der Wasser- und Temperaturverteilung im Bauteil während dem Trocknungsvorgang ist nur sehr schwer möglich und mit großem Aufwand verbunden. Zur Ermöglichung eines detaillierten Einblicks in den zeitlichen Verlauf aller relevanten Prozessgrößen wird ein Simulationsmodell entwickelt.
Die Anwendung von Simulation zur Trocknung von Textilien mittels CFD durch das Lee-Modell [4] wurde in [5] untersucht. Der Trocknungsprozess eines Fleischstücks wurde in [6] anhand einer einphasigen numerischen Strömungssimulation abgebildet. Die Trocknung von Fliesen im industriellen Maßstab wurde in [7] ebenfalls durch eine CFD-Simulation beschrieben. Aus dem Fachgebiet der technischen Keramik sind Publikationen, welche die physikalischen Eigenschaften des Schlickers und die Modellierung eines Trocknungsprozesses über die Finite-Elemente-Methode betreffen, bekannt [8], [9].

Diese Studie befasst sich mit der Modellierung des schichtweisen Trocknungsprozesses von Aluminiumoxidschlicker, wobei ein modifizierter CFD Solver für konjugierte Wärmeübertragungsprobleme zur Anwendung kommt. Die Verdunstung des Wassers und die thermische Kopplung zwischen Schlicker und Luft wird durch eine Randbedingung modelliert. Die physikalischen Eigenschaften des Schlickers werden teilweise aus bekannter Literatur [8] übernommen.
Das Simulationsmodell einer Trocknungsvorrichtung wird erstellt und dessen korrekte Funktionsweise, durch den Vergleich von gemessenen Massen- und Temperaturverläufen mit Simulationsergebnissen, überprüft. Abschließend wird die Validierung und der Vergleich des Simulationsmodells mit bereits zuvor bekannten Lösungen diskutiert. Der hier präsentierte Simulationsansatz weist Vorteile in der Detailreiche im Vergleich mit FE-Simualtionen [9] und geringere Simualtionszeiten als Ansätze, welche auf Mehrphasensimulation basieren, [5] auf.

II. Methoden

Zur Simulation der Vorgänge im Schlicker werden die physikalischen Eigenschaften desselben modelliert und aus Literaturquellen übernommen [8], [9]. Diese werden als Funktion vom Wassermassenanteil Y_W gemeinsam mit einer zusätzlichen Diffusionsgleichung in den OpenFOAM Solver *chtMultiRegionFoam* implementiert. Die Kopplung der Wärme- und Massenübertragung erfolgt über eine Randbedingung.
Der modifizierte Solver und die Randbedingung werden zur Simulation einer Trocknungsvorrichtung angewandt. Zur experimentellen Validierung werden verschiedene Sensoren an derselben positioniert und ausgewertet.

A. Stoffeigenschaften des Schlickers

In dieser Studie wird zur Anfertigung des Schlickers ein thermisch reaktives Aluminiumoxidpulver mit einer Korngrößenverteilung $d90 = 2\,\mu\mathrm{m}$ und einer Dichte von $\rho = 2200\,\mathrm{kg\,m^{-3}}$verwendet. Die prozentuale Zusammensetzung kann Tabelle I entnommen werden.

Tabelle I
ZUSAMMENSETZUNG DES ALUMINIUMOXIDSCHLICKERS MIT DAZUGEHÖRIGEN MASSENANTEILEN UND DICHTE

Bestandteil	Masse /%	Dichte $\rho / \frac{\mathrm{kg}}{\mathrm{m}^{-3}}$
Wasser	21.31	468.82
Aluminiumoxid	77.5	1705
Additive	1.19	26.18

Zur Simulation der Wärmeleitung im als Festkörper modellierten Schlicker ist es notwendig die Dichte des Aluminiumoxid-Wassergemisches ρ in Abhängigkeit des Wassermassenanteils Y_W in $\mathrm{kg\,kg^{-1}}$ zu beschreiben. Dies erfolgt über die Dichte des Aluminiumoxidpulvers ρ_{Al} bei Vernachlässigung von Schrumpfung nach Gleichung (1).

$$\rho = \rho_{Al}\left(1 + \frac{Y_W}{1 - Y_W}\right) \tag{1}$$

Die Wärmekapazität des Schlickers c_p in $\mathrm{J\,kg^{-1}\,K^{-1}}$ errechnet sich ebenso über den Anteil der Komponenten am Gemisch und deren Wärmekapazitäten c_{Al}, c_W nach Gleichung (2).

$$c_p = Y_W(c_W - c_{Al}) + c_{Al} \tag{2}$$

Der Verlauf der Wärmeleitfähigkeit λ in $\mathrm{W\,m^{-1}\,K^{-1}}$, siehe Abbildung 1, wird durch Parameterabgleich eines Polynoms dritter Ordnung mit Messwerten eines vergleichbaren Schlickers aus [8] modelliert.

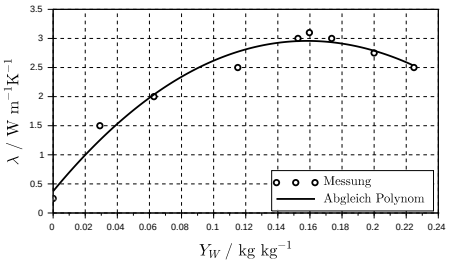


Abbildung 1. Anpassung eines Polynoms dritter Ordnung an die Messdaten der Wärmeleitfähigkeit λ durch Parameterabgleich

Die Koeffizienten des abgeglichenen Polynoms sind Gleichung (3) zu entnehmen.

$$\lambda_p = 30.53\,Y_W^3 - 112.79\,Y_W^2 + 33.41\,Y_W + 0.37 \tag{3}$$

Der Verlauf der Diffusivität von Wasser in Aluminumoxid D_Y in $\mathrm{m^2\,s^{-1}}$ ist aus vorausgehender Forschung [9] bekannt und Gleichung (4) zu entnehmen.

$$D_Y = D_2 + \frac{1}{2}\left(1 + erf\left(\frac{Y_W - Y_{W,tr}}{dY_W\sqrt{2}}\right)\right)(D_1 - D_2) \tag{4}$$

Der Verlauf wird in einen exponentiellen Bereich D_2 und einen konstanten Bereich D_1 unterteilt, wobei die Verbindung derselben über die Gaußsche Fehlerfunktion erf erfolgt. Diese stellt eine Sigmoidfunktion dar, welche beim Massenanteil $Y_{W,tr}$ über eine Intervallbreite von dY_W zwischen exponentiellem und konstantem Bereich wechselt.

B. Solveranpassung

Zur Modellierung des schichtweisen Trocknungsprozesses wird der OpenFOAM Solver *chtMultiRegionFoam* modifiziert. Die massenanteilsabhängigen Stoffeigenschaften, sowie eine zusätzliche Diffusionsgleichung im Festkörper, werden implementiert. Die Simulation von konjugierter Wärme- und Massenübertragung zwischen mehreren Fluid- und

Festkörperdomänen läuft dabei über die Lösungsstrategie aus Abbildung 2 ab.

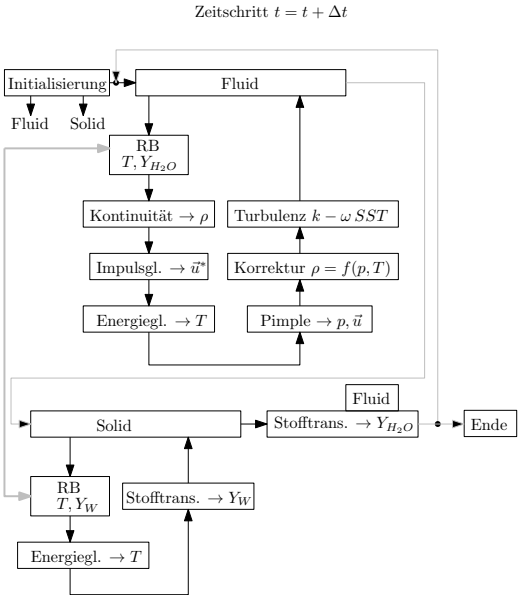


Abbildung 2. Lösungsstrategie des für Trocknungsprozesse adaptierten Solvers *chmtMultiRegionFoam*

Der Solver arbeitet im Fluid nach einem druckbasierten Prinzip und koppelt Wärme- und Massentransport über eine Randbedingung zwischen den Domänen. Die Turbulenzmodellierung erfolgt über das $k\,\omega$-SST Modell, da wandnahe und freie Strömungsbereiche vorhanden sind. Der Transport des Massenbruchs an Wasserdampf Y_{H_2O} erfolgt am Ende des Zeitschritts. Die Zeitintegration wird durch ein explizites Verfahren mit variabler Zeitschrittweitensteuerung über die Courantzahl umgesetzt.
Die Kontinuitätsgleichung ist in kompressibler Form für die Fluiddomäne -wie in Gl.(5) beschrieben- implementiert,

$$\frac{\partial \rho}{\partial t} + \nabla \cdot \rho \vec{u} = 0 \tag{5}$$

wobei ρ der Dichte und $\vec{u}$ dem Geschwindigkeitsfeld in $\mathrm{m\,s^{-1}}$ entspricht. Die Impulserhaltung ist durch die kompressible Form der Navier-Stokes Gleichung, wie in Gleichung (6) dargestellt, implementiert.

$$\frac{\partial \rho \vec{u}}{\partial t} + \nabla \cdot (\rho \vec{u} \vec{u}) = -\nabla p + (\mu + \mu_t) \nabla^2 \vec{u} + \vec{F_b} \tag{6}$$

Die natürliche Konvektion wird durch den Auftriebsterm $\vec{F_b}$ berücksichtigt, wobei p dem Druck in Pa entspricht. Die Summe aus natürlicher Viskosität μ und turbulenter Viskosität μ_t in $\mathrm{m^2\,s^{-1}}$ entspricht der Diffusivität des Impulses.
Die Energiegleichung ist in Form der totalen spezifischen Enthalpie h in $\mathrm{J\,kg^{-1}}$ aus Gleichung (7) implementiert. Dabei wird die kinetische Energie auch mitberücksichtigt.

$$\frac{\partial \rho h}{\partial t} + \nabla \cdot (\rho \vec{u} h) + \frac{\partial \rho (\frac{\vec{u}^2}{2})}{\partial t} + \nabla \cdot \frac{\vec{u}^2}{2} = \alpha_{eff} \nabla^2 h + \rho \vec{u} \cdot \vec{g} + \dot{Q}_{rad} \tag{7}$$

Die kinetische Energie entspricht dem Term $\frac{\vec{u}^2}{2}$ und die effektive thermische Diffusivität α_{eff} in $\mathrm{m^2\,s^{-1}}$ berücksichtigt sowohl die natürliche als auch die turbulente Diffusivität. Die mechanische Energie wird durch den Term $\rho \vec{u} \cdot \vec{g}$ berücksichtigt und $\dot{Q}_{rad}$ entspricht dem Strahlungswärmestrom. Der Transport des Massenanteils an Wasserdampf Y_{H_2O} in $\mathrm{kg\,kg^{-1}}$ wird durch die Skalartransportgleichung (8) modelliert.

$$\frac{\partial \rho Y_{H_2O}}{\partial t} + \nabla \cdot (\rho \vec{u} Y_{H_2O}) = \rho \alpha_v \nabla^2 Y_{H_2O} \tag{8}$$

Der Koeffizient α_v in $\mathrm{m^2\,s^{-1}}$ beinhaltet sowohl die natürliche als auch die turbulente Diffusivität.
Im Festkörper werden die implementierten Diffusionsgleichungen unter Verwendung der vom Wassermassenanteil Y_W abhängigen Stoffeigenschaften gelöst. Die Energiegleichung ist dabei über die Temperatur T in K nach Gleichung (9) implementiert.

$$\frac{\partial}{\partial t} T = \nabla^2 \frac{\lambda}{c_p \rho} T \tag{9}$$

Die Diffusion des Wassers wird über Gleichung (10) abgebildet.

$$\frac{\partial}{\partial t} Y_W = D_Y \nabla^2 Y_W \tag{10}$$

Vor dem Lösen der Transportgleichungen im Festkörper werden die Skalarfelder der jeweiligen Diffusionskoeffizienten anhand der Lösungen aus dem vorigen Zeitschritt berechnet.

C. Kopplungsrandbedingung

Zur Kopplung von Wärme und Massenübertragung zwischen Fluid- und Festkörperdomäne wird eine Randbedingung für die Temperatur T, und die Massenanteile Y_{H_2O} und Y_W erarbeitet. In diesem Simulationsmodell wird zur Vereinfachung angenommen, dass die Verdunstung des Wassers nur an der Kommunikationsfläche zwischen Fluid und Festkörper vonstatten geht. Die Wärmekopplung basiert auf den Gleichungen(11) und (12).

$$\dot{q}_s = -\dot{q}_f + \dot{m} h_{lat} \tag{11}$$

$$T_{ff} = T_{fs} = T_f \tag{12}$$

Es wird Kontinuität im Wärmestrom des Fluids $\dot{q}_f$ und des Festkörpers $\dot{q}_s$ in $\mathrm{W\,m^{-2}}$ vorausgesetzt, wobei der latente Wärmestrom sich aus dem flächenspezifischen Massenstrom $\dot{m}$ in $\mathrm{kg\,m^{-2}s^{-1}}$ und der Verdampfungsenthalpie h_{lat} in $\mathrm{J\,kg^{-1}}$ zusammensetzt. Die Flächentemperaturen T_f in K müssen sowohl für die Fluid- T_{ff}, als auch für die Festkörperdomäne T_{fs} dieselben sein.
Der Massenaustausch wird über die Erhaltungs- und Gleichgewichtsbedingungen nach den Gleichungen (13) und (14) definiert.

$$\frac{\dot{m}_s}{A} = -\frac{\dot{m}_f}{A} \tag{13}$$

$$a_W = RH \tag{14}$$

Kontinuität im Massenstrom zwischen Festkörper $\dot{m}_s$ und Fluid $\dot{m}_f$ in $\mathrm{kg\,m^{-2}s^{-1}}$ werden vorausgesetzt. Gleichgewicht zwischen Wasseraktivität a_W an der Festkörperoberfläche und der relative Luftfeuchte RH muss auch eingehalten werden, wobei die Wasseraktivität durch empirische Daten aus [8], [9] bekannt ist.

Die Berechnungen in der Randbedingung folgen dem Ablauf aus Abbildung 3.

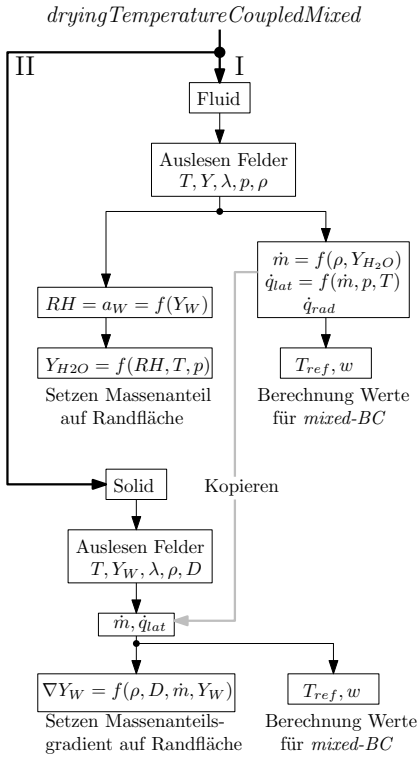


Abbildung 3. Ablauf der Berechnungen in der Randbedingung *dryingTemperatureCoupledMixed* für beide Simulationsdomänen

Die Randbedingung wird, wie in Abbildung 2 ersichtlich, vor dem Lösen der jeweiligen Transportgleichungen zuerst in der Fluid- und anschließend in der Festkörperdomäne aufgerufen. Davon abhängig werden der Massenanteil an Wasserdampf im Fluid oder der Gradient der Wassermassenanteils im Festkörper berechnet und gesetzt. Die Flächentemperatur wird über eine *mixed* Randbedingung umgesetzt.

D. Simulationsmodell

Zur Durchführung einer Simulation mit dem adaptierten Solver und der Randbedingung wird das Modell der Trocknungvorrichtung aus Abbildung 4 erstellt.

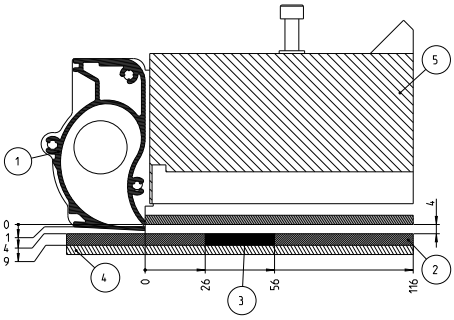


Abbildung 4. Abmessungen und Hauptkomponenten der Trocknungsvorrichtung in Schnittdarstellung

Die Auflistung und Beschreibung der Komponenten ist in Tab.II ausgeführt.

Tabelle II
KOMPONENTEN AUS ABB.4

Nummer:	Bezeichnung:
1	Luftdüse
2	Gummimatte
3	Schlickerriegel
4	Aluminiumplatte
5	Heizer (ausgeschaltet)

Die Austrittsgeschwindigkeit aus der Luftdüse wird als konstant angenommen, diese Annahme ist durch zuvor durchgeführte Messungen überprüft worden. Die Simulationsgeometrie wird in eine Festkörper- und eine Fluidomäne mit den dazugehörigen Randflächen wie in Abbildung 5 dargestellt, unterteilt.

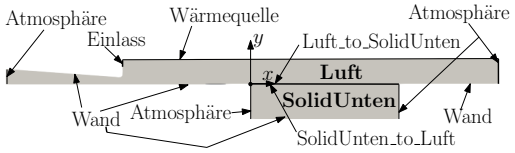


Abbildung 5. Domänen und Randflächen des CFD Modells der Trocknungsvorrichtung

Die Simulation wird zur Reduktion der Rechenzeit in 2D aufgebaut, wobei sich der Einfluss dieser Vereinfachung durch Überprüfung mit einem vollständigen Modell als minimal erwiesen hat. Die in der Luftdomäne gesetzten Randbedingungen und die Initialisierung der Felder sind in Tab.III aufgelistet.

Tabelle III
VORGEGEBENE WERTE DER RAND- UND ANFANGSBEDINGUNGEN IN DER LUFTDOMÄNE

	$u/\frac{\mathrm{m}}{\mathrm{s}}$	T/K	p/Pa	$Y_{H_2O}/\frac{\mathrm{kg}}{\mathrm{kg}}$
Einlass	2	298.45	$95 \cdot 10^4$	0.011
Atmosphäre	PIOV	296.75	$95 \cdot 10^4$	0.011
Wand	0	296.75	$95 \cdot 10^4$	$\nabla Y = 0$
Luft_to_SolidU	0	B	$95 \cdot 10^4$	B
Internes Feld	0	296.75	$95 \cdot 10^4$	0.011

Die in Tabelle III verwendete Abkürzung B steht für Berechnet und PIOV steht für eine *pressureInletOutletVelocity* Randbedingung. In der Soliddomäne werden die Bedingungen aus Tab.IV aufgebracht.

Tabelle IV
VORGEGEBENE WERTE DER RAND- UND ANFANGSBEDINGUNGEN IN DER SOLIDDOMÄNE

	T/K	$Y/\frac{\mathrm{kg}}{\mathrm{kg}}$
SolidUntent_to_Luft	B	B
Wand	296.75	$\nabla Y = 0$
Atmosphäre	296.75	$\nabla Y = 0$
Internes Feld	296.75	0.24

E. Vernetzung

Die Vernetzung wird mit Hexaedern durch den Vernetzungsalgorithmus *snappyHexMesh* realisiert. Der für den Trocknungsvorgang relvante Bereich rund um die Begrenzungsfläche Luft_to_SolidUnten weist ein orthogonales Netz auf. Der Bereich in der Nähe des Einlasses und entlang des Randes der Luftdüse zeigt etwas schlechtere Netzqualität, siehe Abbildung 6.

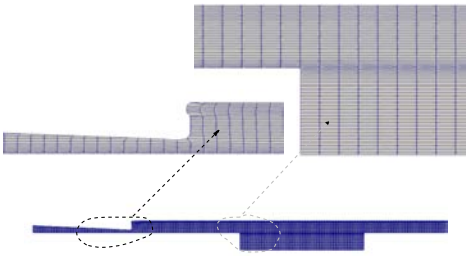


Abbildung 6. Vernetzung an der Kommunikationsfläche zwischen den Domänen und am Auslass der Luftdüse

Die Durchführung einer Qualitätsanlyse des Netzes ergibt die Parameter aus Tabelle V für die jeweilige Domäne.

Tabelle V
AUFLISTUNG DER NETZEIGENSCHAFTEN UND ZELLENANZAHL FÜR BEIDE DOMÄNEN

	Solid	Fluid
Seitenverhältnis	13.26	14.68
max. Nicht-Orthogonalität	2.56	63.73
mittl. Nicht-Orthogonalität	0.075	3.12
Anzahl Hexaeder	13500	39030

Das verwendete Netz ist das Ergebnis einer Konvergenzstudie, wobei zur Reduzierung der Zellenanzahl bei gleicher Ergebnisqualität die Bereiche mit hohen Gradienten feiner aufgelöst sind.

F. Messaufbau zur Modellvalidierung

Zur Validierung des erarbeiteten Simulationsmodells wird ein Abgleich mit experimentell ermittelten Daten durchgeführt. Die Trocknungsvorrichtung wird dazu, wie in Abbildung 7 dargestellt, mit Sensoren beaufschlagt.

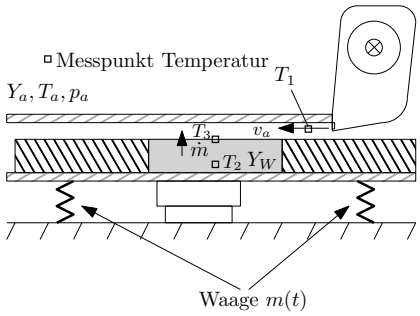


Abbildung 7. Sensorenpositionierung an der Trocknungsvorrichtung zur Validierung des Simulationsmodells

Die prozessrelevanten Messgrößen werden in zwei Arten unterteilt, stationäre bzw. konstante Größen und zeitabhängige Größen. Zu den stationären Größen gehören:

- Massenanteil an Wasserdampf in der Umgebungsluft Y_a und Umgebungstemperatur T_a
- Austrittsgeschwindigkeit an der Luftdüse v_a
- Temperatur T_1 der einströmenden Luft
- Umgebungsdruck $p_a \rightarrow$ relevant für Verdampfung

Als zeitlich veränderlich werden die folgenden Messgrößen betrachtet:

- Masse des Grünlings $m(t)$ über Waage gemessen
- Temperaturen T_2, T_3

Die Temperaturmessung erfolgt über zwei Thermoelemente Typ J (T_1, T_2) und das Strahlungpyrometer Micro-Epsilon CS-SF15-C1 (T_3). Die Luftgeschwindigkeit wird über das Heizdrahtanemometer TESTO 425 und die Masse über die Präzisionswaage Kern 572 gemessen. Die Auswertung der Daten erfolgt über die Software LabView und das NATIONAL INSTRUMENT Chassis NI cDAQ-9188XT, die Waage wird dabei direkt mit dem PC über eine RS232 Schnittstelle verbunden.

III. Ergebnisse

A. Stationäre Messungen an der Trocknungsvorrichtung

Zu Beginn der Messung werden Umgebungsbedingungen, Schlickereigenschaften, Luftgeschwindigkeit und die Festkörper-Abmessungen gemessen und in Tabelle VI dokumentiert.

Tabelle VI
Stationäre und konstante Messgrössen als Eingangsparameter für das Simulationsmodell

Stationäre Messgrößen	
T_a/K	296.75
$RH/\%$	59
$Y_a/\%$	0.01134
p_a/Pa	95000
$v_a/\mathrm{m\,s^{-1}}$	2
T_a/K	298.45
$\rho_i/\mathrm{kg\,m^{-3}}$	2340
$\xi/\%$	76
l_g/mm	90
h_g/mm	4
b_g/mm	30

Die Bezeichnung der Messwerte erfolgt großteils nach Abbildung 7, wobei ξ der Masse des Festkörperanteils in $\%$ im frischen Schlicker und ρ_i der anfänglichen Dichte desselben entspricht. Die Abmessungen des Schlickerriegels werden mit l_g, h_g, b_g bezeichnet.

B. Vergleich zwischen Simulation und Messung

Die stationären Messwerte dienen als Eingangsgrößen für das CFD-Modell zur Simulation des Massenstroms und des Temperaturverlaufs. Zur Vereinfachung werden diese Größen über die gesamte Prozessdauer als konstant angenommen, die Umgebungstemperatur T_a wird zudem auf allen Wänden als *fixedValue*-Randbedingung beaufschlagt.
Die Simulation wird für eine Dauer von $t = 170\,\mathrm{s}$ durchgeführt und mit den aufgezeichneten Messergebnissen verglichen, wie für die relative Massenabnahme dm in Abbildung 8 dargestellt ist.

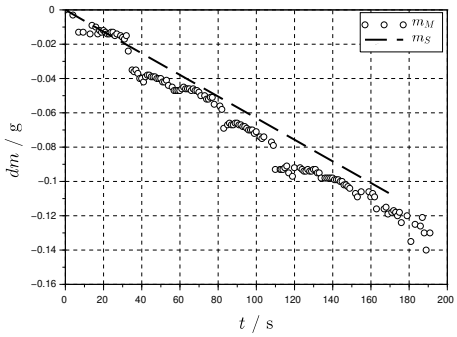


Abbildung 8. Vergleich der Messergebnisse m_M mit den Simualtionsergebnissen m_S bezogen auf die relative Massenabnahme über die Zeit

Zusätzlich zur Massenabnahme ist der Verlauf der Temperaturen an den Messpunkten aus Abbildung 7 zur Charakterisierung des Trocknungsverhaltens interessant, welcher in Abbildung 9 dargestellt ist. Der Index $_M$ steht dabei für Messdaten und $_S$ für Simulationsdaten.

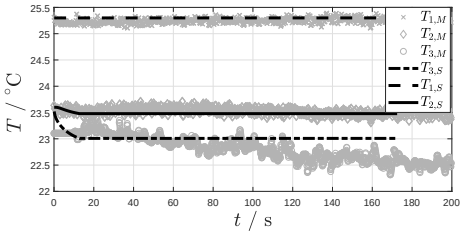


Abbildung 9. Abgleich der Temperaturverläufe von Messung und Simulation über die Zeit

Betrachtet man den Massenverlauf aus Abbildung 8, so ist eine gute Übereinstimmung zwischen Messwerten und Simulation erkennbar. Vorhandene Störungen im Messverlauf sind durch aerodynamische Kräfte aufgrund der Strömung und weitere Umgebungseinflüsse zu begründen, deren Einfluss aufgrund der geringen Massenabnahme ins

Gewicht fällt. Das Vorhandensein eines linearen Verlaufs stimmt auch mit den Ergebnissen aus [8], [9] überein, was durch die hohe Wasseraktivtät $a_W \approx 1$ begründet werden kann. Der Prozess nimmt daher einen stationären Charakter an, bis durch die beginnende Abnahme der Wasseraktivität die zweite Trocknungsphase eingeläutet wird [10].

Der Temperaturverlauf zeigt auch eine gute Übereinstimmung zwischen Mess- und Simulationsergebnissen bezogen auf T_1 und T_2. Die Temperaturabnahme in den Simulationsergebnissen zu Beginn bei T_2 ist durch den Zeitversatz zwischen Simulations- und Messergebnissen, welcher durch das Füllen der Kavität mit Schlicker zwangsläufig auftritt, zu begründen. Zu Beginn der Messaufzeichnung ist bereits ein quasi-stationärer Zustand erreicht.

Eine größere Abweichung zwischen Simulation und Messung ist am Verlauf von T_3 erkennbar, wobei der zu Beginn vorhandene Temperaturabfall in den Simulationsergebnissen wieder durch den Zeitversatz der Messungen begründet werden kann. Der über den gesamten Zeitbereich gemessene Temperaturabfall ist in den Simulationsergebnissen nicht vorhanden, dies kann durch die idealisierten Randbedingungen und die daher fehlende Abkühlung der Gummiumrandung begründet werden.

IV. Diskussion

Es wurde erfolgreich gezeigt, dass es mittels CFD-Simulation möglich ist, die physikalischen Abläufe bei schichtweisen Trocknungsprozessen abzubilden. Getroffene Vereinfachungen wie die ortsfeste Verdunstung an der Kommunikationsfläche und die Vernachlässigung der Schrumpfung wurden evaluiert und anhand der guten Übereinstimmung mit den Messergebnissen legitimiert. Auch ohne die Verwendung von Mehrphasenmodellen wie in [5] konnten bei geringerem Rechenaufwand alle relevanten physikalischen Vorgänge realitätsnah simuliert werden. Dabei konnte der Rechenaufwand durch die Verwendung einer *mixed* im Vergleich zu [6] leicht reduziert werden.

Auch im Vergleich mit Lösungsansätzen, welche auf der Finite-Elemente-Methode basieren [9],[8], zeigte das entwickelte Simulationmodell Vorteile in der Genauigkeit. Dazu gehört die lokale Berücksichtigung der strömungsbedingten Konvektion auf der Kommunikationsfläche und die dadurch mögliche Bestimmung von lokalen Problemstellen in der Trocknung. Den logischen nächsten Schritt stellt die Anwendung des Simulationsmodells in der System- und Prozessentwicklung dar.

Ein Beispiel dafür wäre die Durchführung von Parameterstudien zur Bestimmung des Einflusses der verschiedenen Prozessgrößen auf das Trocknungsergebnis und die Evaluierung neuer Trocknergeometrien. Jene Erkenntnisse, welche durch Anwendung des Modells zukünftig generiert werden können, stellen in Kombination mit begleitenden Messungen den Grundstein für eine effiziente Prozess- und Maschinenentwicklung dar.

Danksagung

Ich möchte mich beim MCI Innsbruck, der Firma D3-AM GmbH und der Firma Durst Group AG für die Unterstützung bei der Umsetzung dieser Forschungsarbeit bedanken. Ein weiterer Dank geht an József Nagy von eulerian-solutions e.U. für die Einschulung in OpenFOAM und die Unterstützung bei Implementierungsproblemen.

Literatur

[1] U. Scheithauer, E. Schwarzer, T. Moritz, and A. Michaelis, "Additive manufacturing of ceramic heat exchanger: opportunities and limits of the lithography-based ceramic manufacturing (lcm)," *Journal of Materials Engineering and Performance*, vol. 27, no. 1, pp. 14–20, 2018.

[2] M. Schwentenwein, P. Schneider, and J. Homa, "Lithography-based ceramic manufacturing: A novel technique for additive manufacturing of high-performance ceramics," in *Advances in Science and Technology*, vol. 88. Trans Tech Publ, 2014, pp. 60–64.

[3] E. Willems, M. Turon-Vinas, B. C. dos Santos, B. Van Hooreweder, F. Zhang, B. Van Meerbeek, and J. Vleugels, "Additive manufacturing of zirconia ceramics by material jetting," *Journal of the European Ceramic Society*, vol. 41, no. 10, pp. 5292–5306, 2021.
[4] W. Lee, "A pressure iteration scheme for two-phase modeling. technical report la-ur 79-975," *Los Alamos, New Mexico*, 1979.
[5] T. Blejchař, J. Raška, and J. Jablonská, "Mathematical simulation of drying process of fibrous material," in *EPJ Web of Conferences*, vol. 180. EDP Sciences, 2018, p. 02010.
[6] F. Trujillo and Q. Pham, "Cfd modeling of heat and moisture on a two dimensional model of a beef leg," in *International Congress of Refrigeration*, 2003.
[7] W. Kriaa, S. Bejaoui, H. Mhiri, G. Le Palec, and P. Bournot, "Study of dynamic structure and heat and mass transfer of a vertical ceramic tiles dryer using cfd simulations," *Heat and Mass Transfer*, vol. 50, no. 2, pp. 235–251, 2014.
[8] S. Oummadi, "Drying behaviour of ceramic green bodies: experimental characterization and numerical modelling," Ph.D. dissertation, Université de Limoges, 2019.
[9] N. Lauro, S. Oummadi, A. Alzina, B. Nait-Ali, and D. S. Smith, "Computer model of drying behaviour of ceramic green bodies with particular reference to moisture content dependent properties," *Journal of the European Ceramic Society*, vol. 41, no. 14, pp. 7321–7329, 2021.
[10] P. Erhard, J. Angenoorth, J. Vogt, J. Spiegel, F. Ettemeyer, W. Volk, and D. Günther, "Characterization of slurry-cast layer compounds for 3d printing of high strength casting cores," *Materials*, vol. 14, no. 20, p. 6149, 2021.

Oliver Ungerer studiert am Mechatronik Department am MCI Innsbruck/ Österreich und arbeitet bei D3-AM GmbH in der Maschinen- und Prozessentwicklung.