

Konferenzband 2023 | Proceedings 2023

Masterarbeiten | Master Theses

Brocksieper, Jonas	Design Analysis of a Model-Based Rotor Position Observer for Traction Motors	17
Deiser, Simon	Admittance Control and Adaptive Assistance Rehabilitation for a Portable Hand Rehabilitation Device	33
Eder, Simon	Model Predictive Control of an Electrodynamic Shaker	41
Federer, Daniel	Produktionsoptimierung von Defibrillationselektroden mithilfe einer roboterunterstützten Teilautomatisierung	49
Felbermayr, Dimitri	Periodic boundary conditions for non-uniform meshes	57
Geiger, Dominique Mathäus	Data Structure for Virtual Commissioning - Exchange Relevant Information Focusing on Integration of Modular Component Models and Wide Compatibility	65
Gfall, Christian	An Educational System for Advanced Control Engineering: Development of an Intelligent Self-Balancing Robot	73
Hofmeister, Gregor	Experimental determination of the material parameters of an alpine ski	82
Höglauer Johannes	Development of a Vacuum-tight Exchange Mechanism for Mass Spectrometers	90
Huber, Dominik	Ausbildungsplattform Robotik - Regelung von Industrierobotern	98

Kofler, Michael	Entwicklung eines Impulshammers für Schienensysteme	106
Lederer, Thomas	Intelligenter Entmistungsroboter für die Reinigung von Spaltenböden	112
Leps, Dominik	Entwicklung eines Messsystems zur Erfassung der spektralen UV-Intensitäten an einem Quecksilberdampf-Strahler	121
Ludwig, Dominic	Photogrammetrie des Olympia Eiskanals Innsbruck und Implementierung in ein virtuelles Erlebnis	129
Martino, Daniel	Concept Development for the Rotor Assembly Line of an Interior Permanent Magnet Synchronous Motor (IPMSM)	143
Mikula, Florian	Erweiterung der experimentellen Bestimmung der Biegesteifigkeit eines Alpinksis	151
Mütterlein, Florian	Implementation of monocular depth cue in 3D Software by the means of motion parallax and eye-tracking	159
Prantner, Jan	Drone-Based 3D Mapping and Virtual Reality Inspection for Sport Facilities	167
Reinmüller, Stefan	Optimierung der Erkennungsrate von Machine Learning Algorithmen zur Detektion von Defekten in der Zahnproduktion	173
Ruppert, Patrick	Methoden zur Durchführung Accelerated-Life-Time-Tests am Beispiel eines Rasenmähers	181
Sanner, Johan Cornelis	Developing a gas engine emissions model for controls validation	190
Seewald, Sebastian	Development of a Portable Hand Rehabilitation Device for Sensory Motor Training	198
Strobel, Hansi	Deep Learning-based Condition Monitoring Framework for Rolling-Element Bearings	206
Triendl, Fabian	Immersive co-located visuo-haptic feedback in 3D space	214

Design Analysis of a Model-Based Rotor Position Observer for Traction Motors

Jonas Brocksieper*, Maurizio Incurvati** (supervisor), and Julius Rogowski* (supervisor)

Abstract—Permanent magnet synchronous machines offer the highest efficiency and power density making them well-suited for automotive drive trains. In order to exploit the motor's full potential, advanced torque control requires knowledge of the rotor position. Position sensors increase costs, are prone to fail and inaccurate at higher speeds. As an alternative, position sensorless control algorithms are increasingly being used. In this paper, a Kalman Filter is used to investigate the accuracy and applicability of four different model-based rotor angle observers. The results show that all models are suitable for mid- and high-speed operation. The estimation errors differ at lower speeds. The models suited for lower speeds require higher computational effort. Using the estimated angle for the motor control does not lead to additional instabilities. Compared to the previously installed resolver, the rotor angle observer partly leads to reduced torque deviations which indicates a higher angle accuracy. The findings described in this paper can be used to determine an appropriate observer type and machine model for model-based rotor angle observation in industrial applications.

Index Terms—PMSM, Sensorless Control, Rotor Position Estimation, Kalman Filter.

I. INTRODUCTION

In an attempt to bring range, cost, weight and reliability of the electric drive train to the level of combustion engines, automotive manufacturers are constantly searching for new optimization potentials to exploit. Currently, a permanent magnet synchronous motor (PMSM) as a traction drive delivers the best performance with respect to these criteria. A car is accelerated by the traction motor

exerting torque on the wheels. Controlling the car's motion therefore means controlling the motor torque. Advanced control strategies like field oriented control (FOC) are designed to deliver the requested torque with maximum efficiency, dynamics and accuracy independent of the operating point.

Controlling the torque of an synchronous motor (ASM) requires knowledge of the rotor angle. This information is usually obtained by means of a rotor position sensor. To save costs and increase robustness in electric drives, algorithms that calculate or estimate the rotor position on the basis of known system variables are increasingly being used. These position-sensorless methods either use the back-induced voltage or the saliency of the rotor to estimate the rotor angle.

In recent publications, model-based angle observers have been proposed. They make use of both the back electromotive force (BEMF) and the saliency. In this paper, these observers shall be analyzed by evaluating the applicability of an extended Kalman filter with different underlying machine models. The main question is what machine model is most suitable for traction drives regarding accuracy, robustness against machine parameter tolerances, central processing unit load, integrability and overall performance at low motor speed.

II. FUNDAMENTALS

A. Kalman Filter

The Kalman Filter (KF) is a model-based approach to estimate the unmeasured states and possibly obtain an optimized estimation of the measured states. Both

*Mercedes-Benz AG

**MCI Management Center Innsbruck, Department Mechatronics

the measurement noise \vec{v}_k and the process noise \vec{w}_k can be characterized by a mean and a covariance.

$$\vec{x}_k = \underline{A}_{k-1} \vec{x}_{k-1|k-1} + \underline{B}_{k-1} \vec{u}_k + \vec{w}_k \quad (1)$$

$$\vec{y}_k = \underline{C}_k \vec{x}_k + \vec{v}_k \quad (2)$$

The KF reduces the covariance of the state estimates to a minimum by taking into account both the measurement covariance \underline{R}_k and the process noise covariance \underline{Q}_k . The KF assumes the error mean of both the measurement and the model to be zero.

The Extended Kalman Filter (EKF) is an extension of the standard KF that can handle nonlinear models. The continuous-time differential form of the model is first linearized and subsequently discretized to obtain the discrete-time matrices used in the EKF algorithm. In the prediction step, the nonlinear state transition and measurement models are linearized by computing the Jacobian matrices.

$$\underline{F}_k = \left. \frac{\partial \vec{f}}{\partial \vec{x}} \right|_{\hat{\vec{x}}_{k-1|k-1}} + \left. \frac{\partial (g\vec{u}_k)}{\partial \vec{x}} \right|_{\hat{\vec{x}}_{k-1|k-1}} \quad (3)$$

$$\underline{H}_k = \left. \frac{\partial \vec{h}}{\partial \vec{x}} \right|_{\hat{\vec{x}}_{k|k-1}} \quad (4)$$

The states and error covariance matrix are predicted according to equations 5 and 6. The Kalman gain \underline{K} from equation 7 is used to correct the state prediction (8) and covariance prediction (9).

$$\hat{\vec{x}}_{k|k-1} = \vec{f}(\hat{\vec{x}}_{k-1|k-1}) + g(\hat{\vec{x}}_{k-1|k-1})\vec{u}_k \quad (5)$$

$$\underline{P}_{k|k-1} = \underline{F}_k \underline{P}_{k-1|k-1} \underline{F}_k^T + \underline{Q}_k \quad (6)$$

$$\underline{K}_k = \underline{P}_{k|k-1} \underline{H}_k^T \left(\underline{H}_k \underline{P}_{k|k-1} \underline{H}_k^T + \underline{R}_k \right)^{-1} \quad (7)$$

$$\hat{\vec{x}}_{k|k} = \hat{\vec{x}}_{k|k-1} + \underline{K}_k \left(\vec{y}_k - \vec{h}(\hat{\vec{x}}_{k|k-1}) \right) \quad (8)$$

$$\underline{P}_{k|k} = (\underline{I} - \underline{K}_k \underline{H}_k) \underline{P}_{k|k-1} \quad (9)$$

III. IPMSM MODEL EQUATIONS

The magnetic flux vector $\vec{\Psi}_{dq}$ of an interior-buried permanent magnet synchronous Motor (IPMSM) is a function of the current vector \vec{i}_{dq} .

$$\begin{bmatrix} \Psi_d \\ \Psi_q \end{bmatrix} = \begin{bmatrix} L_d & 0 \\ 0 & L_q \end{bmatrix} \begin{bmatrix} i_d \\ i_q \end{bmatrix} + \begin{bmatrix} \Psi_{PM} \\ 0 \end{bmatrix} \quad (10)$$

The current vector $\vec{i}_{\alpha\beta}$ can be described as a function of the voltage vector $\vec{u}_{\alpha\beta}$ using Ohm's law and the induction law.

$$\begin{bmatrix} u_\alpha \\ u_\beta \end{bmatrix} = R \begin{bmatrix} i_\alpha \\ i_\beta \end{bmatrix} + \frac{d}{dt} \begin{bmatrix} \Psi_\alpha \\ \Psi_\beta \end{bmatrix} \quad (11)$$

Transforming 11 to the dq frame, replacing $\vec{\Psi}_{dq}$ with 10 and subsequent rearrangement yields:

$$\frac{d}{dt} \begin{bmatrix} i_d \\ i_q \end{bmatrix} = \begin{bmatrix} \frac{1}{L_d} & 0 \\ 0 & \frac{1}{L_q} \end{bmatrix} \begin{bmatrix} u_d \\ u_q \end{bmatrix} + \begin{bmatrix} -\frac{R}{L_d} & \frac{\omega L_q}{L_d} \\ -\frac{\omega L_d}{L_q} & -\frac{R}{L_q} \end{bmatrix} \begin{bmatrix} i_d \\ i_q \end{bmatrix} - \frac{\omega \Psi_{PM}}{L_q} \begin{bmatrix} 0 \\ 1 \end{bmatrix} \quad (12)$$

Highly exploited machines, e.g. automotive traction motors, exhibit nonlinear relationships between $\vec{\Psi}_{dq}$ and \vec{i}_{dq} . With rising \vec{i}_{dq} , the change rate of $\vec{\Psi}_{dq}$ decreases. This effect is called magnetic flux saturation. Figure 1 shows Ψ_d and Ψ_q as functions of both i_d and i_q for an exemplary saturating machine.

The flux time derivative $\frac{d\vec{\Psi}_{dq}}{dt}$ can be split into the partial differential of $\vec{\Psi}_{dq}$ after the \vec{i}_{dq} and the time derivative of \vec{i}_{dq} [1]:

$$\begin{aligned} \frac{d\vec{\Psi}_{dq}}{dt} &= \frac{\partial \vec{\Psi}_{dq}}{\partial \vec{i}_{dq}} \frac{d\vec{i}_{dq}}{dt} = \begin{bmatrix} \frac{\partial \Psi_d}{\partial i_d} & \frac{\partial \Psi_d}{\partial i_q} \\ \frac{\partial \Psi_q}{\partial i_d} & \frac{\partial \Psi_q}{\partial i_q} \end{bmatrix} \frac{d}{dt} \begin{bmatrix} i_d \\ i_q \end{bmatrix} \\ &= \begin{bmatrix} L_{dd} & L_{dq} \\ L_{qd} & L_{qq} \end{bmatrix} \frac{d}{dt} \begin{bmatrix} i_d \\ i_q \end{bmatrix} \end{aligned} \quad (13)$$

L_{dd} , L_{dq} , L_{qd} and L_{qq} are the so-called differential inductances in contrast to the absolute inductances L_d and L_q . If, in good approximation, the

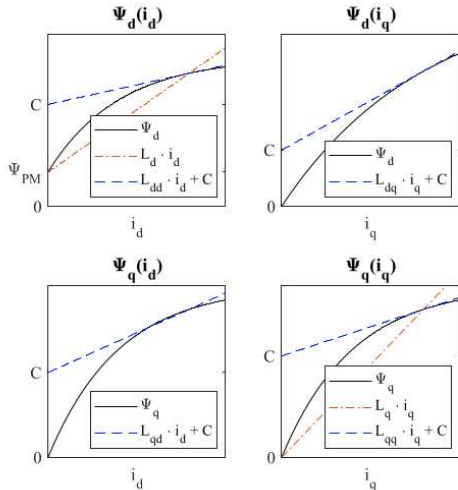


Figure 1. Saturation of Ψ_d and Ψ_q as functions of i_d and i_q

magnetic flux is proportional to the current, absolute and differential inductances are the same:

$$\begin{aligned} L_{dd} &= L_d & L_{dq} &= 0 \\ L_{qq} &= L_q & L_{qd} &= 0 \end{aligned} \quad (14)$$

With a real machine we additionally find that i_d also has an influence on Ψ_q and also i_q on Ψ_d . This phenomenon is called cross saturation and represented by the inductances L_{dq} and L_{qd} . Cross saturation is not to confuse with cross coupling where Ψ_d induces a speed-dependent u_q and vice versa.

Thus we obtain the nonlinear machine model equation:

$$\begin{aligned} \begin{bmatrix} u_d \\ u_q \end{bmatrix} &= \begin{bmatrix} R & -\omega L_q \\ \omega L_d & R \end{bmatrix} \begin{bmatrix} i_d \\ i_q \end{bmatrix} + \omega \Psi_{PM} \begin{bmatrix} 0 \\ 1 \end{bmatrix} \\ &+ \begin{bmatrix} L_{dd} & L_{dq} \\ L_{qd} & L_{qq} \end{bmatrix} \frac{d}{dt} \begin{bmatrix} i_d \\ i_q \end{bmatrix} \end{aligned} \quad (15)$$

Both \vec{i}_{dq} and its time derivative appear in equations

12 and 15 making it them differential equations. Digital controllers are only able to execute iterative algorithms. In order to achieve high model accuracy in the discrete-time domain, the fourth order Runge–Kutta method (RK4) [2] is used to solve the model equations in discrete time.

Linear systems can be discretized analytically. The inverse Laplace transform method (equation 16) is chosen to discretize the linear models. From this point onwards, continuous-time matrices will be indicated by a \sim over the symbol, e.g. $\tilde{\underline{A}}$.

$$\underline{A} = \mathcal{L}^{-1} \left\{ \left[\underline{I}s - \tilde{\underline{A}} \right]^{-1} \right\} \quad (16)$$

with $\tilde{\underline{A}}$ being the diagonal matrix of the eigenvalues of \underline{A} and \underline{P} the matrix of the corresponding eigenvectors. The input and output matrices are discretized as follows:

$$\underline{B} = \int_0^T \underline{A}(\tau) d\tau \tilde{\underline{B}} \quad (17)$$

$$\underline{C} = \tilde{\underline{C}} \quad (18)$$

IV. MODEL-BASED ROTOR ANGLE ESTIMATION

In this paper, four different machine models are used to observe the rotor angle. They are suitable for angle observation either based on the principle of model reconstruction or else by including the motor dynamics.

A. Rotor Angle Estimation Based on Model Reconstruction

If the rotor rotates, a vector that is stationary in the dq frame is visible as two sinusoidal signals in the $\alpha\beta$ frame. If the vector is fixed to a known axis, the \arctan of the ratio between the vector amplitudes along the α and β axis determines the angle of this axis in the $\alpha\beta$ frame. The goal of model reconstruction is rearrange the PMSM equations to find a quantity that is fixed to either the d or the q axis.

Both the Extended Flux model [3] and Extended EMF model [4] apply this strategy to determine the rotor angle. The first develops the concept of the extended flux Ψ_{ext} which points along the d axis in stationary conditions. Besides developing an axis fixed quantity, this reconstruction leads to a model formulation in the $\alpha\beta$ system with reduced parameter dependency.

The second model develops the concept of the extended electromotive force (EMF) e_{ext} which always points along the q axis. Both models use their respective axis-fixed quantities as states \vec{x} additional to the stator current vector $\vec{i}_{\alpha\beta}$. The stator currents also form the model output \vec{y} while the stator voltages are the model input \vec{u} .

$$\begin{aligned} \vec{y} &= \begin{bmatrix} i_\alpha \\ i_\beta \end{bmatrix} & \vec{u} &= \begin{bmatrix} u_\alpha \\ u_\beta \end{bmatrix} \\ \vec{x}_{Flux} &= \begin{bmatrix} i_\alpha \\ i_\beta \\ \Psi_{ext,\alpha} \\ \Psi_{ext,\beta} \end{bmatrix} & \vec{x}_{EMF} &= \begin{bmatrix} i_\alpha \\ i_\beta \\ e_{ext,\alpha} \\ e_{ext,\beta} \end{bmatrix} \end{aligned} \quad (19)$$

B. Rotor Angle Estimation Including the Motor Dynamics

The electrical equations of a salient machine transformed into the $\alpha\beta$ system contain the mechanical states θ and ω . By including an additional mechanical input like the load torque T_L , the mechanical equations for position and speed can be included in the model.

$$\begin{aligned} \vec{y} &= \begin{bmatrix} i_\alpha \\ i_\beta \end{bmatrix} & \vec{u} &= \begin{bmatrix} u_\alpha \\ u_\beta \\ T_L \end{bmatrix} \\ \vec{x}_{Combined} &= \begin{bmatrix} i_d \\ i_q \\ \omega \\ \theta \end{bmatrix} & \vec{x}_{Stator} &= \begin{bmatrix} i_\alpha \\ i_\beta \\ \omega \\ \theta \end{bmatrix} \end{aligned} \quad (20)$$

The differential equation for ω is developed from the torque equilibrium equation

$$\begin{aligned} \frac{d\omega}{dt} &= \alpha = \frac{p}{J}T_{EM} - \frac{p}{J}T_L - \frac{B_f}{J}\omega \\ &= \frac{3p^2}{2J}(\Psi_{PM} + L_\Delta i_d)i_q - \frac{p}{J}T_L - \frac{B_f}{J}\omega \end{aligned} \quad (21)$$

The current equations of the Combined model are taken from equation 12 by transforming the voltage vector into the $\alpha\beta$ system:

$$\begin{aligned} \frac{d}{dt} \begin{bmatrix} i_d \\ i_q \end{bmatrix} &= \begin{bmatrix} \frac{\cos\theta}{L_d} & \frac{\sin\theta}{L_d} \\ -\frac{\sin\theta}{L_q} & \frac{\cos\theta}{L_q} \end{bmatrix} \begin{bmatrix} u_\alpha \\ u_\beta \end{bmatrix} \\ &+ \begin{bmatrix} -\frac{R}{L_d} & \frac{\omega L_q}{L_d} \\ -\frac{\omega L_d}{L_q} & -\frac{R}{L_q} \end{bmatrix} \begin{bmatrix} i_d \\ i_q \end{bmatrix} - \frac{\omega \Psi_{PM}}{L_q} \begin{bmatrix} 0 \\ 1 \end{bmatrix} \end{aligned} \quad (22)$$

The Stator model formulation is developed in [5].

C. Observability

The observability condition of all four mentioned models has been developed in [6]. The Extended Flux and Extended EMF model are observable for $\omega \neq 0$. The Combined and Stator model are observable at $\omega = 0$ if $\frac{di_q}{dt} \neq 0$ and $L_d \neq L_q$.

V. MACHINE MODEL CORRECTION TO INCLUDE MAGNETIC SATURATION

The models have to be adapted in order to account for nonlinear magnetic flux and cross saturation.

The following definitions are used to simplify model notation:

$$\begin{aligned} L_\Sigma &= L_d + L_q & L_{\Sigma dd} &= L_{dd} + L_{qq} & L_{\Sigma dq} &= L_{dq} + L_{qd} \\ L_\Delta &= L_d - L_q & L_{\Delta dd} &= L_{dd} - L_{qq} & L_{\Delta dq} &= L_{dq} - L_{qd} \\ L_{2\pi} &= 2L_d L_q & L_{2\pi dd} &= 2L_{dd} L_{qq} & L_{2\pi dq} &= 2L_{dq} L_{qd} \end{aligned}$$

A. Nonlinear Extended Flux Model

With L_q being a time-dependent quantity, the product rule has to be applied when differentiating $\Psi_{\alpha\beta}$. From the resulting equation, the system matrix

$\tilde{\underline{A}}$ and input matrix $\tilde{\underline{B}}$ are drawn. $\tilde{\underline{B}}$ stays the same compared to the linear formulation described in [3].

$$\tilde{\underline{A}} = \begin{bmatrix} -\frac{R+\frac{d}{dt}L_q}{L_q} & 0 & 0 & \frac{\omega}{L_q} \\ 0 & -\frac{R+\frac{d}{dt}L_q}{L_q} & -\frac{\omega}{L_q} & 0 \\ 0 & 0 & 0 & -\omega \\ 0 & 0 & \omega & 0 \end{bmatrix} \\ = \begin{bmatrix} a & 0 & 0 & c \\ 0 & a & -c & 0 \\ 0 & 0 & 0 & -\omega \\ 0 & 0 & \omega & 0 \end{bmatrix} \quad (23)$$

To simplify the analytic discretization of 23, the matrix elements are assigned to the variables a and c .

$$\underline{A} = \mathcal{L}^{-1} \left\{ \left[Ls - \tilde{\underline{A}} \right]^{-1} \right\} \\ = \mathcal{L}^{-1} \left\{ \begin{bmatrix} \frac{1}{s-a} & 0 & \frac{c}{s-a} \frac{\omega}{s^2+\omega^2} & \frac{c}{s-a} \frac{s}{s^2+\omega^2} \\ 0 & \frac{1}{s-a} & -\frac{c}{s-a} \frac{\omega}{s^2+\omega^2} & \frac{c}{s-a} \frac{\omega}{s^2+\omega^2} \\ 0 & 0 & \frac{s}{s^2+\omega^2} & -\frac{\omega}{s^2+\omega^2} \\ 0 & 0 & \frac{\omega}{s^2+\omega^2} & \frac{s}{s^2+\omega^2} \end{bmatrix} \right\} \\ = \begin{bmatrix} e^{aT} & 0 & k(T) & l(T) \\ 0 & e^{aT} & -l(T) & k(T) \\ 0 & 0 & \cos \omega T & -\sin \omega T \\ 0 & 0 & \sin \omega T & \cos \omega T \end{bmatrix} \quad (24)$$

with

$$k(T) = \frac{c}{a^2 + \omega^2} \left(\omega \left(e^{aT} - \cos \omega T \right) - a \sin \omega T \right) \quad (25)$$

$$l(T) = \frac{c}{a^2 + \omega^2} \left(a \left(e^{aT} - \cos \omega T \right) + \omega \sin \omega T \right) \quad (26)$$

$k(T)$ and $l(T)$ result from the multiplication of two Laplace domain functions. According to the convolution theorem, a multiplication of two functions in the Laplace domain is the same as the convolution of the according time-dependent functions and vice versa:

$$\mathcal{L}^{-1} \left\{ F(s)G(s) \right\} = f(t) * g(t) \quad (27)$$

The discretized input matrix \underline{B} becomes

$$\underline{B} = \int_0^T \underline{A}(\tau) d\tau \tilde{\underline{B}} = \frac{e^{aT}}{aL_q} \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \quad (28)$$

B. Nonlinear Extended EMF Model

Considering the nonlinear magnetic flux, the Extended EMF model is reformulated as follows:

$$\begin{bmatrix} u_d \\ u_q \end{bmatrix} = \begin{bmatrix} R & -\omega L_q \\ \omega L_q & R \end{bmatrix} \begin{bmatrix} i_d \\ i_q \end{bmatrix} \\ + \begin{bmatrix} L_{dd} & L_{dq} \\ -L_{dq} & L_{dd} \end{bmatrix} \frac{d}{dt} \begin{bmatrix} i_d \\ i_q \end{bmatrix} + \begin{bmatrix} 0 \\ e_{ext} \end{bmatrix} \quad (29)$$

For the sake of completeness, the extended EMF e_{ext} changes to:

$$e_{ext} = \omega(L_{\Delta}i_q + \Psi_{PM}) - (L_{dq} + L_{qd}) \frac{di_d}{dt} - L_{\Delta dd} \frac{di_q}{dt} \quad (30)$$

Transformation into the stator frame yields:

$$\begin{bmatrix} u_{\alpha} \\ u_{\beta} \end{bmatrix} = \begin{bmatrix} R - \omega L_{dq} & \omega(L_{dd} - L_q) \\ \omega(L_q - L_{dd}) & R - \omega L_{dq} \end{bmatrix} \begin{bmatrix} i_{\alpha} \\ i_{\beta} \end{bmatrix} \\ + \begin{bmatrix} L_{dd} & L_{dq} \\ -L_{dq} & L_{dd} \end{bmatrix} \frac{d}{dt} \begin{bmatrix} i_{\alpha} \\ i_{\beta} \end{bmatrix} + \begin{bmatrix} e_{ext,\alpha} \\ e_{ext,\beta} \end{bmatrix} \quad (31)$$

The system matrix $\tilde{\underline{A}}$ and input matrix $\tilde{\underline{B}}$ can be represented by

$$\tilde{\underline{A}} = \begin{bmatrix} a & -b & -c & d \\ b & a & -d & -c \\ 0 & 0 & 0 & -\omega \\ 0 & 0 & \omega & 0 \end{bmatrix}, \quad \tilde{\underline{B}} = \begin{bmatrix} c & -d \\ d & c \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \quad (32)$$

with

$$\begin{aligned} a &= \frac{1}{m}(\omega L_q L_{dq} - L_{dd} R) \\ b &= \frac{1}{m}(\omega(m - L_q L_{dd}) + L_{dq} R) \\ c &= \frac{L_{dd}}{m} \\ d &= \frac{L_{dq}}{m} \\ m &= L_{dd}^2 + L_{dq}^2 \end{aligned} \quad (33)$$

The discretized system matrix \underline{A} becomes

$$\underline{A} = \begin{bmatrix} e^{aT} \cos bT & -e^{aT} \sin bT & k(T) & l(T) \\ e^{aT} \sin bT & e^{aT} \cos bT & -l(T) & k(T) \\ 0 & 0 & \cos \omega T & -\sin \omega T \\ 0 & 0 & \sin \omega T & \cos \omega T \end{bmatrix} \quad (34)$$

Due to their complexity, $k(T)$ and $l(T)$ are not discretized analytically but approximated by using the Taylor expansion of second order:

$$k(T) = \frac{T^2}{2}(bd - ac + d\omega) - Tc \quad (35)$$

$$l(T) = \frac{T^2}{2}(ad + bc + c\omega) + Td \quad (36)$$

The discretized input matrix \underline{B} becomes

$$\underline{B} = \begin{bmatrix} cp(T) - dq(T) & -cq(T) - dp(T) \\ cq(T) + dp(T) & -cp(T) + dq(T) \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \quad (37)$$

with $p(T)$ and $q(T)$ resulting from the integral

$$\begin{aligned} p(T) &= \int_0^T e^{a\tau} \cos b\tau \, d\tau \\ &= \frac{1}{a^2 + \omega^2} \left(e^{aT} (a \cos bT + b \sin bT) - a \right) \end{aligned} \quad (38)$$

respectively

$$\begin{aligned} q(T) &= \int_0^T e^{a\tau} \sin b\tau \, d\tau \\ &= \frac{1}{a^2 + \omega^2} \left(e^{aT} (a \sin bT - b \cos bT) + b \right) \end{aligned} \quad (39)$$

C. Nonlinear Combined Model and Nonlinear Stator Model

The width of one column is not sufficient to display the full model equations of both the Combined model and the Stator model. They can be found in appendices A and B.

Due to the high nonlinearity of both models, the fourth order Runge-Kutta method is applied to achieve sufficient accuracy.

In order to avoid a torque estimator, the angular acceleration α is taken as the third model input instead of the load torque T_L . α is taken from the last time-step to reduce the model complexity. For fast transients this results in a slight phase lag of the observer, but the accuracy loss is not noticeable so that α can even be filtered heavily. This further improves robustness. The observability at stand-still is not affected since the current equations remain the same.

D. Filter

A phase-locked loop (PLL) is applied to the estimated angles of all models in order to increase the observer stability while maintaining the observer dynamics. A block diagram of the implemented version of this filter is presented in figure 2. It resembles that proposed by [7] with a PT1 filter applied to

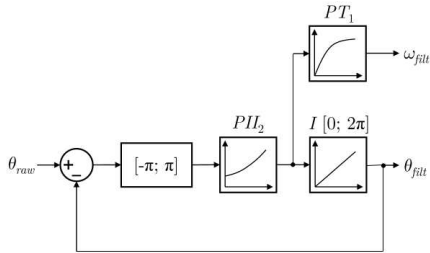


Figure 2. Angle filter with PII_2 controller to follow the rotor acceleration without phase lag

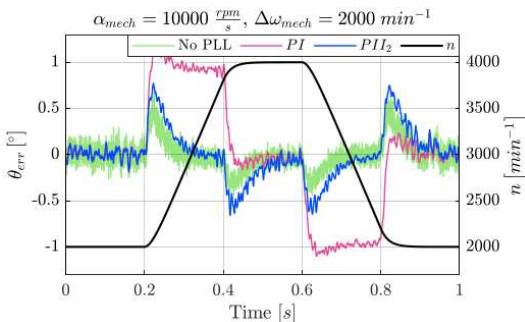


Figure 3. Angle Filter with and without acceleration compensation at 200 Nm

ω since the speed changes slowly compared to the angle. However, the PI controller normally used to drive the phase lag to zero is replaced by a PII_2 controller. This avoids a stationary angle error during acceleration phases.

The green trace in figure 3 shows the Stator model angle error in the simulation without any filter for four different combinations of acceleration and speed. The angle error traces of the PI controller (pink) leaves a stationary angle offset during acceleration phases. The PII_2 controller (blue) however allows the phase-locked loop (PLL) to follow θ_{raw} during $\alpha \neq 0$ without stationary angle error.

VI. OPEN LOOP PERFORMANCE

In a first step, the performance of the different rotor angle observers is analyzed without using the estimated angle for motor control. All measurements were taken on a test bench with a motor from a series-production car at a rotor temperature of 60 °C.

A. Accuracy

The observer accuracy can be best evaluated by analyzing the stationary estimation error throughout the entire operating space. In order to be usable in the traction drives of electric cars, the angle error has to be within $\pm 2^\circ$.

The error mean of the implemented models is shown in figure 4. The Combined and Stator model show the almost exact same accuracy characteristics. In comparison, the Extended EMF model and Extended Flux model slightly differ in accuracy but are roughly comparable. At lower speeds however, the mechanical models are a bit more accurate.

For all models, the angle error with positive torque is very close to zero. At roughly -150 Nm all models show relatively high errors of up to -3° . This is caused by inaccurate motor parameters for these operating points.

B. Stability

Not only the observer accuracy but also stability is an important criterion. In order to evaluate the stability, the signal-to-noise ratio (SNR) of the observed angle is analyzed. The SNR can be characterized by the standard deviation of the time signal.

The angle error takes the encoder signal as a reference for the true rotor angle. Encoder signals however often exhibit noise themselves. The used resolver signal also was oscillating with the first harmonic of the rotor speed. Therefore the angle error includes noise which does not come from the observer. In order to reduce external noise as much as possible, the first harmonic induced by the encoder is filtered out. This is done by first identifying angle and amplitude of the harmonic via the Direct

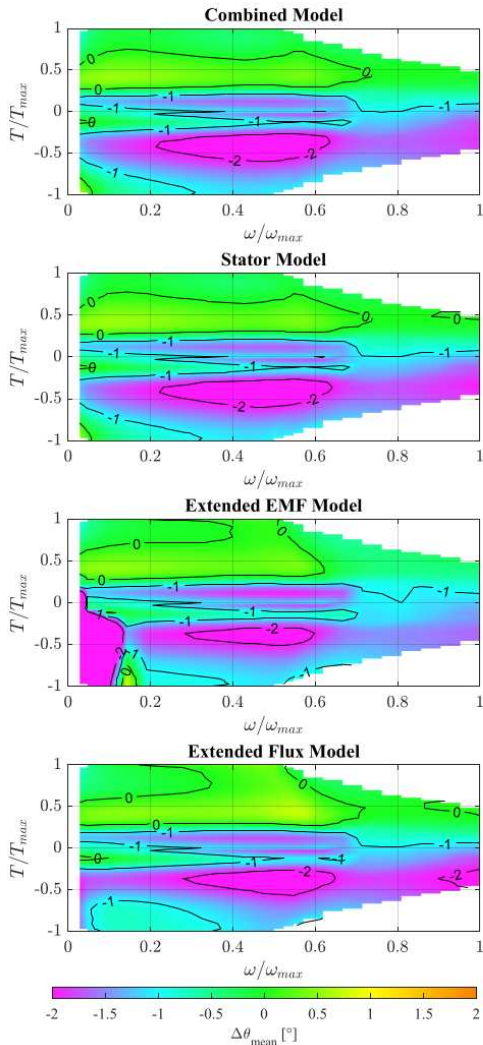


Figure 4. Mean of angle error for each model

Fourier Transformation (DFT) and then subtracting this oscillation from the signal.

Figure 5 shows the root-mean-square (RMS) value of the estimated angle for the entire torque-speed-plane of the Combined model and the Stator model. Both models show high stability in all points. Only at low speeds and negative torque does the Stator model oscillate at higher amplitude. This asymmetry again can be attributed to the parameter inaccuracies which affect the stator model noise more than the extended flux model.

The two models using axis-fixed quantities (Extended Flux and Extended EMF) both become unstable at low speeds. This confirms the results of the observability analysis.

C. Effect of Different Discretization Methods

Finding the right discretization method is often a balance of accuracy and computational effort. With linear systems, the analytical discretization offers the best accuracy which might come at high computational costs. The Taylor approximation is only marginally less accurate but cheaper on the computational side, especially for models consisting of multiple differential equations.

For both the Extended EMF and Extended Flux model, numerical discretization using the Taylor series of second order yields similar performance compared to the analytical discretization. Higher polynomial orders of the Taylor expansion doesn't make a noticeable difference anymore compared to second order.

At a certain numerical accuracy, a further improvement does not benefit the overall accuracy since the noise floor coming from the measurements is higher. The uncertainty of the system parameters going into the system model at one points also becomes bigger than the numerical uncertainty. In short: The limiting factor is often the system knowledge, not the computing power.

For nonlinear models, model discretization with Runge-Kutta higher than fourth order doesn't make a difference. Using the Heun or even Euler method

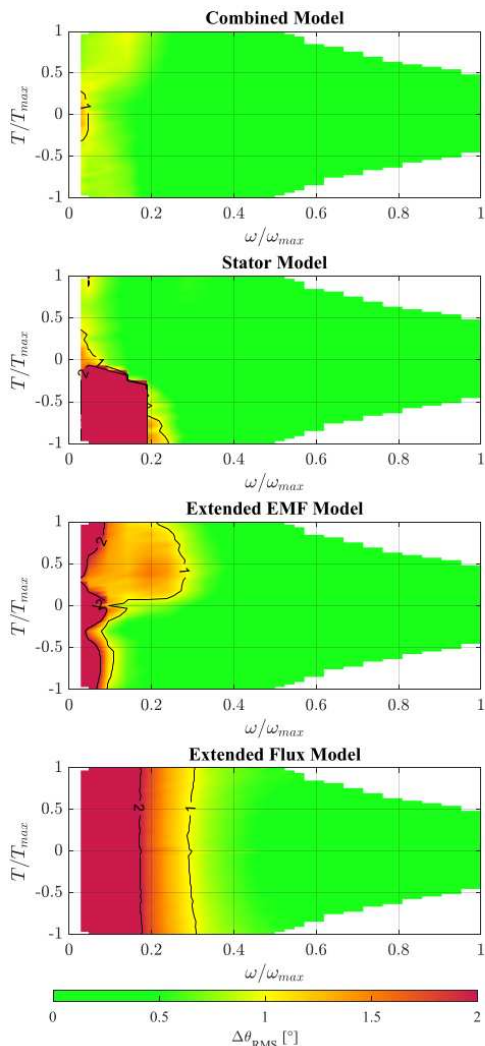


Figure 5. RMS value of angle error for each model

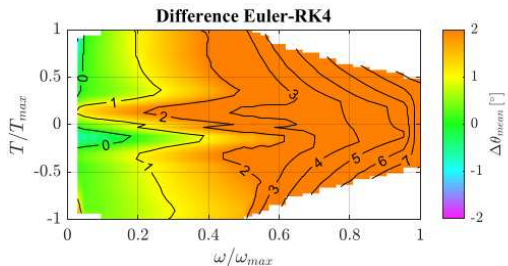


Figure 6. Angle error proportional to motor speed with Euler method

decreases the accuracy noticeably. As shown in figure 6, the Euler method leads to error magnitudes well beyond the requirements.

VII. CLOSED LOOP PERFORMANCE

In order to validate that the motor can be controlled with the developed angle observers, the observers have to be tested in closed loop with the controller. Closed-loop operation means that the estimated rotor angle is fed back to the controller instead of resolver angle. In order to quantify the impact on the motor operation, we compare the angle error in open-loop and closed-loop operation against each other.

The effect on observer accuracy can be seen in figure 7. The angle error mean of both the Combined model and the Stator stays mostly the same. Only at maximum speed does the angle error increase marginally in closed-loop operation. 0.2° however doesn't affect the torque control much.

The same can be said for the observer stability in closed-loop. The differences again are below 0.5° as shown in figure 8. Closed-loop operation therefore doesn't have a considerable effect on the observer stability.

The most realistic scenario on a test bench is created by taking the measured torque and speed traces from a drive and using them as real-time set points for the test bench. The input data of the tested driving cycle covers primarily the medium- and high-speed region of the electric motor and includes highly

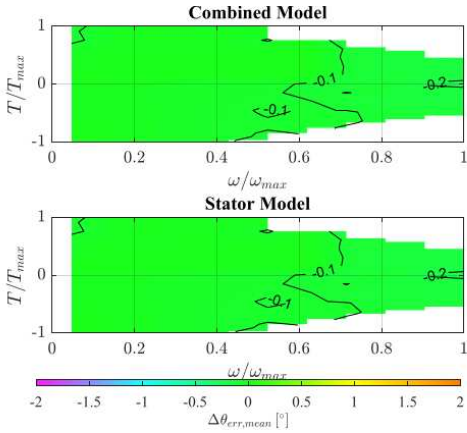


Figure 7. Difference of angle error mean between open and closed loop

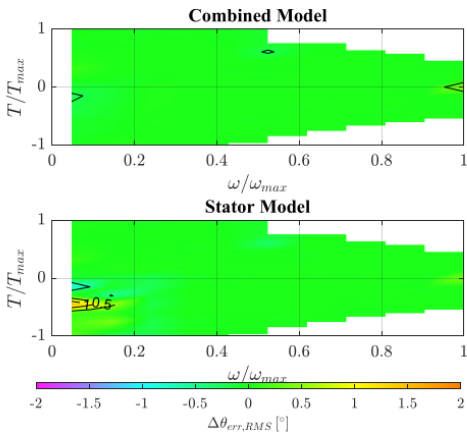


Figure 8. Difference in angle error RMS value between open and closed loop with the Combined model and Stator model

dynamic sections including gear changes. Since the driving profile only includes short passages at low speeds, it is not suited for evaluation of the low-speed observer performance. During the tests, the rotor temperature was kept at around 60 °C. This

ensures better comparability between the observers. The results of this test are presented in this chapter.

The test was conducted with all machine models in closed-loop as well as once in open-loop to create a reference. Each observer was then individually tested in closed-loop. To evaluate the performance of each observer, the difference in torque error between the resolver and the observers is calculated.

The relative torque error is calculated after the following formula:

$$T_{err} = \frac{T_{ref} - T}{\max\left(\frac{T_{max}}{2}, T_{ref}\right)} \quad (40)$$

At $T_{ref} < \frac{T_{max}}{2}$ the relative torque error is referred to $\frac{T_{max}}{2}$ instead of T_{ref} in order to avoid division by 0. The results are presented in figure 9.

The color of each operating point indicates whether the torque accuracy using the estimated angle is better (blue), equal (green) or worse (red) than the torque accuracy using the resolver angle.

Interestingly, at medium speeds and $T < 0$, where figure 4 shows higher angle errors, the torque error is similar if not smaller than with the resolver angle. That points to resolver inaccuracies.

The most blue dots at $|T_{EM}| \gg 0$ are found in the Stator model plot. Visually, there are many turquoise and blue colored dots. Simultaneously there are only few yellow and orange dots. This indicates that the Stator model generally improves torque accuracy compared to the resolver angle.

Again the Combined model is slightly worse than its counterpart. It is worth pointing out that along the T_{max} line in the field-weakening region, the Combined model reaches higher torque accuracy than any other observer. The Extended EMF model and Extended Flux model in comparison lead rather to an increase of the torque error, especially at medium speed and $T_{EM} > 0$. The torque error of the Extended EMF and Flux model becomes relatively big particularly at the corner speed where field-weakening starts reducing T_{max} .

All observers achieve a considerable increase in torque accuracy when regenerating energy ($T_{EM} <$

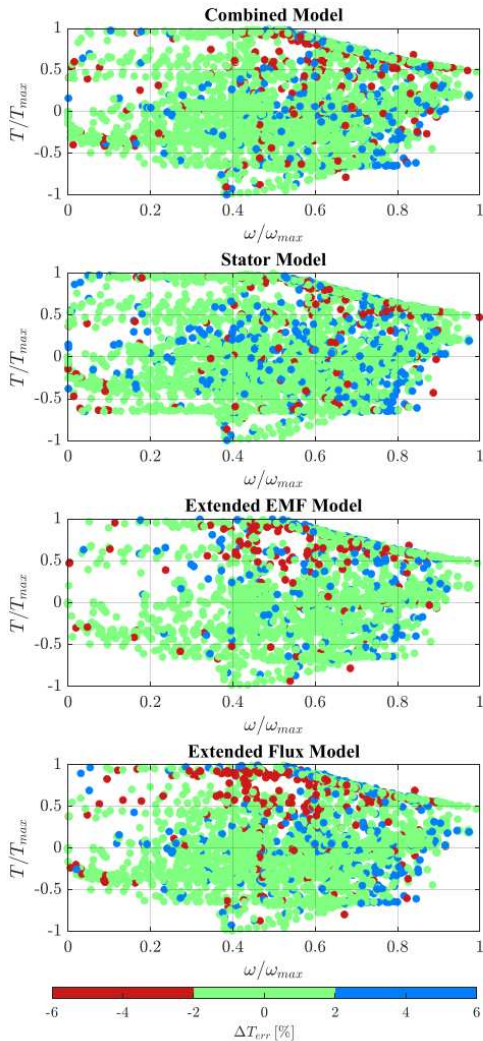


Figure 9. Mean of angle error for each model

0). This is especially the case in the field-weakening area. In order to quantify the overall performance of the resolver and the observers in the driving cycle, the average torque error is calculated and visualized in figure 10.

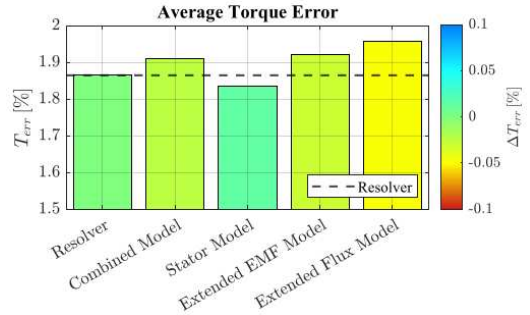


Figure 10. Comparison of average torque error of resolver and observers

The average torque error of each angle observer is on a similar level compared to current control with the resolver. This reflects the analysis results of the previous figures.

To sum up the performance, scores are assigned to each observer for the different criteria shown in table I. Since high-frequency signal injection (HFI) has not been tested on the Combined and Stator model, their standstill performance cannot be evaluated.

Table I
EVALUATION OF THE CONSIDERED STATE SPACE MODELS

	Combined	Stator	Ext. EMF	Ext. Flux
θ_{err} at $\omega = 0$?	?	0	0
θ_{err} at $ \omega > 0$	8	9	6	4
θ_{err} at $ \omega \gg 0$	9	10	8	8
θ_{err} at fast transients	5	6	3	4
Parameter sensitivity	2	3	7	9
Computational effort	3	1	8	10

VIII. CONCLUSION

The test bench measurements have shown that all introduced models are suitable for rotor angle observation at medium and high speeds. The observer accuracy and stability however strongly depend on adequate discretization methods since the numerical error is critical.

Additional filters like a PLL for the angle further improve observer stability. Although low-speed observability is typically problematic for angle observers, both the Combined model and the Stator model have shown sufficient accuracy and stability even at lowest speeds down to zero. The Extended EMF model and the Extended Flux model on the other hand are not observable at low speeds and therefore not suitable for sensorless control of traction drives.

In order to finally evaluate the accuracy at zero speed, the Stator model and Combined model have to be tested with high-frequency voltage injection.

REFERENCES

- [1]
- [2] B. Zeigler, A. Muzy, and E. Kofman, *Theory of Modeling and Simulation*. Amsterdam: Elsevier Inc., 2018.
- [3] Y. Zhao, Z. Zhang, W. Qiao, and L. Wu, "An Extended Flux Model-Based Rotor Position Estimator for Sensorless Control of Salient-Pole Permanent-Magnet Synchronous Machines," *IEEE Transactions on Power Electronics*, vol. 30, no. 8, pp. 4412–4422, 2015.
- [4] A. Eilenberger and M. Schroedl, "Extended Back EMF Model for PMSM with Different Inductances in d and q Axis," pp. 945–948, 2008.
- [5] M. Boussak, "Implementation and Experimental Investigation of Sensorless Speed Control with Initial Rotor Position Estimation for Interior Permanent Magnet Synchronous Motor Drive," *IEEE Transactions on Power Electronics*, vol. 20, no. 6, pp. 1413–1422, 2005.
- [6] R. Nelles and J. Bonifacio, "Model Assessment for Sensorless Control of IPMSM in Automotive Applications," 2021.
- [7] J. T. Friedmann, "Reduktion des Sensorikbedarfs in der industriellen Antriebstechnik," Ph.D. dissertation, Technische Universität München, 2016.

LIST OF FIGURES

1	Saturation of Ψ_d and Ψ_q as functions of i_d and i_q	3
2	Angle filter with PII_2 controller to follow the rotor acceleration without phase lag	7
3	Angle Filter with and without acceleration compensation at 200 Nm	7
4	Mean of angle error for each model	8
5	RMS value of angle error for each model	9
6	Angle error proportional to motor speed with Euler method	9
7	Difference of angle error mean between open and closed loop	10
8	Difference in angle error RMS value between open and closed loop with the Combined model and Stator model	10
9	Mean of angle error for each model	11
10	Comparison of average torque error of resolver and observers	11

APPENDIX

A. Nonlinear Combined Model

$$\begin{aligned}
\tilde{\underline{A}} &= \begin{bmatrix} \frac{1}{m}(\omega L_d L_{dq} - L_{qq} R) & \frac{1}{m}(\omega L_q L_{qq} + L_{dq} R) & \frac{1}{m}(L_{dq} \Psi_{PM}) & 0 \\ \frac{1}{m}(-\omega L_d L_{dd} + L_{dq} R) & \frac{1}{m}(-\omega L_q L_{qd} - L_{dd} R) & \frac{1}{m}(-L_{dd} \Psi_{PM}) & 0 \\ 0 & \frac{3p^2}{2J}(\Psi_{PM} + L_{\Delta} i_d) & -\frac{B_f}{J} & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \\
\tilde{\underline{B}} &= \begin{bmatrix} \frac{1}{m}(L_{qq} \cos \theta + L_{dq} \sin \theta) & \frac{1}{m}(L_{qq} \sin \theta - L_{dq} \cos \theta) & 0 \\ -\frac{1}{m}(L_{dd} \sin \theta + L_{qd} \cos \theta) & \frac{1}{m}(L_{dd} \cos \theta - L_{qd} \sin \theta) & 0 \\ 0 & 0 & -\frac{p}{J} \\ 0 & 0 & 0 \end{bmatrix}, \\
\tilde{\underline{C}} &= \begin{bmatrix} \cos \theta & -\sin \theta & 0 & 0 \\ \sin \theta & \cos \theta & 0 & 0 \end{bmatrix},
\end{aligned} \tag{41}$$

with

$$m = L_{dd} L_{qq} - L_{dq} L_{qd} \tag{42}$$

$$\begin{aligned}
\tilde{\underline{F}} &= \begin{bmatrix} f_{11} & f_{12} & f_{13} & f_{14} \\ f_{21} & f_{22} & f_{23} & f_{24} \\ f_{31} & f_{32} & -\frac{B_f}{J} & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \\
\tilde{\underline{H}} &= \begin{bmatrix} \cos \theta & -\sin \theta & 0 & -i_d \sin \theta - i_q \cos \theta \\ \sin \theta & \cos \theta & 0 & i_d \cos \theta - i_q \sin \theta \end{bmatrix}
\end{aligned} \tag{43}$$

with

$$\begin{aligned}
f_{11} &= \frac{\omega L_d L_{dq} - L_{qq} R}{m} \\
f_{12} &= \frac{\omega L_q L_{qq} + L_{dq} R}{m} \\
f_{13} &= \frac{L_{dq}(L_d i_d + \Psi_{PM}) + L_q L_{qq} i_q}{m} \\
f_{14} &= \frac{u_\alpha(-L_{qq} \sin \theta + L_{dq} \cos \theta) + u_\beta(L_{qq} \cos \theta + L_{dq} \sin \theta)}{m} \\
f_{21} &= \frac{-\omega L_d L_{dd} + L_{dq} R}{m} \\
f_{22} &= \frac{-\omega L_q L_{qd} - L_{dd} R}{m} \\
f_{23} &= \frac{-L_{dd}((L_d i_d + \Psi_{PM}) + L_q L_{qd} i_q)}{m} \\
f_{24} &= \frac{u_\alpha(L_{qd} \sin \theta - L_{dd} \cos \theta) - u_\beta(L_{qd} \cos \theta + L_{dd} \sin \theta)}{m} \\
f_{31} &= \frac{3p^2}{2J} L_\Delta i_q \\
f_{32} &= \frac{3p^2}{2J} (\Psi_{PM} + L_\Delta i_d)
\end{aligned} \tag{44}$$

B. Nonlinear Stator Model

$$\tilde{\underline{A}} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & 0 \\ a_{21} & a_{22} & a_{23} & 0 \\ a_{31} & a_{32} & -\frac{B_f}{J} & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \quad \tilde{\underline{B}} = \begin{bmatrix} b_{11} & b_{12} & 0 \\ b_{21} & b_{22} & 0 \\ 0 & 0 & -\frac{p}{J} \\ 0 & 0 & 0 \end{bmatrix} \tag{45}$$

with

$$\begin{aligned}
a_{11} &= \frac{-2R(L_{\Sigma_{dd}} - L_{\Delta_{dd}} \cos 2\theta) + \omega m \sin 2\theta}{2L_{2\pi_{dd}}} \\
a_{12} &= \frac{2RL_{\Delta_{dd}} \sin 2\theta + \omega(L_{\Sigma_{dd}}(L_{\Sigma} - L_{\Sigma_{dd}}) + L_{\Delta_{dd}}(L_{\Delta} + L_{\Delta_{dd}}) - m \cos 2\theta)}{2L_{2\pi_{dd}}} \\
a_{13} &= \frac{\Psi_{PM} \sin \theta}{L_{qq}} \\
a_{21} &= \frac{2RL_{\Delta_{dd}} \sin 2\theta + \omega(L_{\Sigma_{dd}}(L_{\Sigma_{dd}} - L_{\Sigma}) + L_{\Delta_{dd}}(L_{\Delta} + L_{\Delta_{dd}}) - m \cos 2\theta)}{2L_{2\pi_{dd}}} \\
a_{22} &= \frac{-2R(L_{\Sigma_{dd}} + L_{\Delta_{dd}} \cos 2\theta) - \omega m \sin 2\theta}{2L_{2\pi_{dd}}} \\
a_{23} &= -\frac{\Psi_{PM} \cos \theta}{L_{qq}} \\
a_{31} &= -\frac{3p^2 \sin \theta}{2J}(\Psi_{PM} + L_{\Delta}(i_{\alpha} \cos \theta + i_{\beta} \sin \theta)) \\
a_{32} &= \frac{3p^2 \cos \theta}{2J}(\Psi_{PM} + L_{\Delta}(i_{\alpha} \cos \theta + i_{\beta} \sin \theta)) \\
m &= L_{\Sigma}L_{\Delta_{dd}} + L_{\Delta}L_{\Sigma_{dd}}
\end{aligned} \tag{46}$$

and

$$\begin{aligned}
b_{11} &= \frac{1}{L_{2\pi_{dd}}}(L_{\Sigma_{dd}} - L_{\Delta_{dd}} \cos 2\theta) \\
b_{12} &= -\frac{L_{\Delta_{dd}}}{L_{2\pi_{dd}}} \sin 2\theta \\
b_{21} &= b_{12} \\
b_{22} &= \frac{1}{L_{2\pi_{dd}}}(L_{\Sigma_{dd}} + L_{\Delta_{dd}} \cos 2\theta)
\end{aligned} \tag{47}$$

$$\tilde{\underline{F}} = \begin{bmatrix} a_{11} & a_{12} & f_{13} & f_{14} \\ a_{21} & a_{22} & f_{23} & f_{24} \\ f_{31} & f_{32} & -\frac{B_f}{J} & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \tag{48}$$

with

$$\begin{aligned}
f_{13} &= \frac{-mi_{q_2} + i_b(L_{\Sigma_{dd}}(L_{\Sigma} - L_{\Sigma_{dd}}) + L_{\Delta_{dd}}(L_{\Delta} + L_{\Delta_{dd}}))}{2L_{2\pi_{dd}}} + \frac{\Psi_{PM} \sin \theta}{L_{qq}} \\
f_{14} &= \frac{2L_{\Delta_{dd}}(Ri_{q_2} - u_{q_2}) + \omega mi_{d_2}}{L_{2\pi_{dd}}} + \frac{\omega \Psi_{PM} \cos \theta}{L_{qq}} \\
f_{23} &= \frac{-mi_{d_2} - i_{\alpha}(L_{\Sigma_{dd}}(L_{\Sigma_{dd}} - L_{\Sigma}) + L_{\Delta_{dd}}(L_{\Delta} + L_{\Delta_{dd}}))}{2L_{2\pi_{dd}}} - \frac{\Psi_{PM} \cos \theta}{L_{qq}} \\
f_{24} &= \frac{2L_{\Delta_{dd}}(Ri_{d_2} - u_{d_2}) - \omega mi_{q_2}}{L_{2\pi_{dd}}} + \frac{\omega \Psi_{PM} \sin \theta}{L_{qq}} \\
f_{31} &= \frac{3p^2}{2J}(L_{\Delta}i_{q_2} - \Psi_{PM} \sin \theta) \\
f_{32} &= \frac{3p^2}{2J}(L_{\Delta}i_{d_2} + \Psi_{PM} \cos \theta) \\
i_{d_2} &= i_{\alpha} \cos 2\theta + i_{\beta} \sin 2\theta \\
i_{q_2} &= -i_{\alpha} \sin 2\theta + i_{\beta} \cos 2\theta \\
u_{d_2} &= u_{\alpha} \cos 2\theta + u_{\beta} \sin 2\theta \\
u_{q_2} &= -u_{\alpha} \sin 2\theta + u_{\beta} \cos 2\theta
\end{aligned} \tag{49}$$

Admittance Control and Adaptive Assistance Rehabilitation for a Portable Hand Rehabilitation Device

Simon Deiser, Yeongmi Kim (supervisor)

Abstract—Every year about 10 million people suffer from a stroke worldwide and the number is increasing. About 80% of the patients have permanent motoric disabilities where a post therapy is sufficient. Due to that high amount, an efficient home rehabilitation quote is necessary for the future. Hand motor skills in particular require a very intensive therapy in order to re-learn the complex mechanisms and fine motor skills.

The thesis describes the development of a system that is able to measure the closing or opening forces of the fingers/hand and the use of this data within several control strategies. The hardware is build up in a modular way and can be replaced easily. The force measurement is done with force sensing resistor (FSR) sensors that can detect finger closing and opening forces. To use the data of the sensors, four different software programs are developed. An admittance control mode is used to generate a defined resistance against the movement of the user. An impedance controller is implemented to simulate objects with different strengths that can be gripped by the patient. To help users with more severe motoric impairments, a user intention mode is developed. In this mode, the patient has to cross a patient adaptive force threshold once, to open or close the device fully. With the help of the assistance-as-needed mode, the user is assisted only when necessary, once he is no longer able to continue the movement.

The emerged system enables an intuitive interaction with the user. The haptic feedback generated by the admittance and impedance control algorithms give the user the feeling similar to touching objects of different materials. This makes the therapy more realistic and allows to simulate tasks of daily life.

Index Terms—Admittance control, Impedance control, Tele-rehabilitation, adaptive assistance, hand

S. Deiser is with the Department of Mechatronics, MCI, Innsbruck, Austria, e-mail: s.deiser@mci4me.at.

I. INTRODUCTION

THE number of stroke incidents is increasing worldwide, requiring new more efficient solutions for rehabilitation. Within one year about 1.1 Mio. stroke incidents are registered in Europe. It is estimated that the number of stroke prevalent cases will increase about 27% from 2017 to 2047 [1]. Stroke is the leading cause for disabilities and therefore has a significant impact on the public health system [2]. About 85% of the stroke survivors suffer from motoric disabilities of the hand or arm. After 3 to 6 months still 55-75% have limitations of the hand or arm functions [3]. Due to the high number of patients an optimal care is hard to achieve with the conventional therapy methods. Studies show that the needed therapy frequency cannot be achieved, which prolongs the rehabilitation phase [4]. As a result, alternative therapy methods are getting more and more important. Robotic therapy devices can help to increase the therapy frequency and therefore achieve better therapy outcomes [5]. In addition, there is evidence that an early start of therapy has a high effect on the outcome [5]. Unfortunately, the non-portable robotic devices cannot be used for early therapy initiation when the patient is still bedridden. For this reason, there is a high potential for portable handheld rehabilitation devices.

There are already existing devices for hand rehabilitation on the market (SaeboGlove®, Tyromotion AMADEO®, Hocoma ManovoSpring®). However,

the portable devices only have a passive support. All devices with active support are not flexible to use due to the big size and complexity. In addition, the active systems are due to the complexity very expensive and therefore only available in specialised centres / hospitals.

In Figure 1 the previous portable hand rehabilitation device (PHRD) which provides simple hand opening, closing with only preselected ranges and speed.

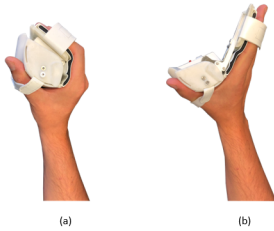


Fig. 1. (a) Portable rehabilitation device for the hand in closed state (b) Opened state

II. METHODS

A. Hardware

By the use of Autodesk Inventor different prototypes are designed. The hardware designed is focused on the door part of the existing device (blue marked part in Figure 2). The CAD parts are printed with an FDM (Fused Deposit Modeling) 3D printer and assembled accordingly.

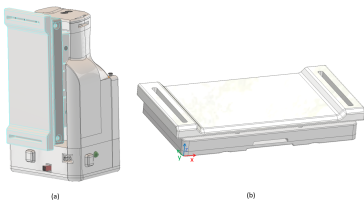


Fig. 2. (a) Assembly of the device with the integrated developed door part (marked blue). (b) Isolated view of the designed door part

To detect the force that is applied by the user's hand, FSR sensors are implemented. Four sensors each are implemented to measure the force in the respective direction. FSR sensors of the type FSR06-BE are used for the concept.

The prototype door part is build up of five main components. The parts are described below:

- (1) - Top cover of the door, where the hand is attached on. The pushing or pulling force is transferred with this part.
- (2) - Contact element between the cover part and the FSR sensors to ensure an optimal use of the active surface of the sensors.
- (3) - FSR sensors.
- (4) - Bottom part that is attached to the opening guidance of the device.
- (5) - Fixation / velcro strap to fix the hand on the door part. The strap is attached on the Top cover (1) so that the pulling force is transferred to the sensors.

The top cover (1) is able to move free in the z-direction to detect the pulling or pushing force of the user's hand. With the fixation (5) the hand is fastened on the cover and the pulling force is transmitted to (1). With the FSR sensors (3) the slight movements of (1) are detected. To concentrate the force on the active surface of the FSR sensors, contact elements made of a rubber material are used. The parts for the prototype are printed with a 3D-printer (FDM - fused deposit modeling).

B. Electronics

The sensors are connected to a printed circuit board (PCB) where the signal is amplified. The amplification is done with a voltage divider.

The output Voltage V_{OUT} can be calculated with the following equation:

$$V_{OUT} = \frac{R_M}{R_M + R_{FSR}} V_{CC} \quad (1)$$

For the setup, a reference input voltage $V_{CC} = 5\text{ V}$ is provided, a FSR sensor of the type FSR406-BE and a pull-down resistor $R_M = 10\text{ K}\Omega$ is used. The electrical components can be placed inside the housing of the developed prototype. The main signal cables are connected to the main electronics board of the device.

The relationship of the applied pushing force and the voltage output is shown in Figure 3.

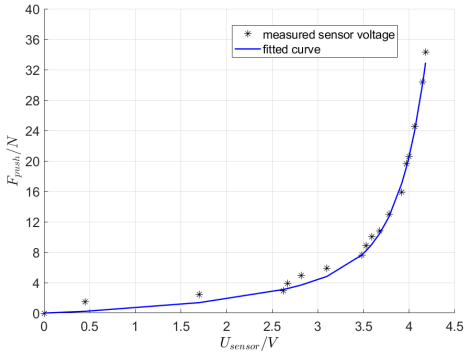


Fig. 3. Relationship between the applied force and the voltage output of the averaged four FSR sensors in PUSH direction

To calculate the force from the voltage output of the averaged FSR sensors, the characteristic curve is fitted with an exponential function with two terms.

$$F_{push} = a \cdot e^{b \cdot U_{sensor}} + c \cdot e^{d \cdot U_{sensor}} - 0.706 \quad (2)$$

With the coefficients $a = 0.7063$, $b = 0.6337$, $c = 7.401 \cdot 10^{-6}$ and $d = 3.583$.

The same procedure is done in the pulling direction. For the fitting function of the pull direction the same equation with different parameters is used.

$$F_{pull} = a \cdot e^{b \cdot U_{sensor}} + c \cdot e^{d \cdot U_{sensor}} + 0.45 \quad (3)$$

With the coefficients $a = 24.16$, $b = 1.006$, $c = -24.61$ and $d = 0.9943$. By the use of the two equations, the forces in both directions can be calculated with regard to the sensor output voltage. Therefore, the equations are implemented in the software.

The measurement in the pushing direction is done with an developed fixture, as can be seen in Figure 4. The measuring points were recorded with different weights between 100g and 3500g. For the pulling direction, the door part is turned 180° .

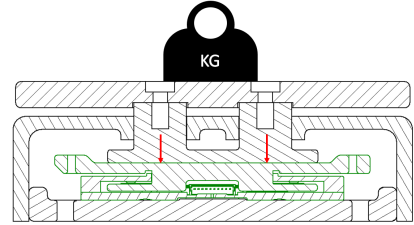


Fig. 4. CAD design of the measurement setup to evaluate the sensor voltage of by applying a pushing force / weight. The green / grey (in black/white print) part shows the designed door part where the force is applied.

C. Software

In this section the different control strategies are described.

1) *Admittance control*: The admittance control uses a defined force to generate a resistance against the movement of the user. The resistive force is again modeled with a spring-damper system to generate a linear viscoelastic system. The stiffness and damping factor can be chosen by the user / therapist or automatically determined by the system. The general idea is shown in Figure 5.

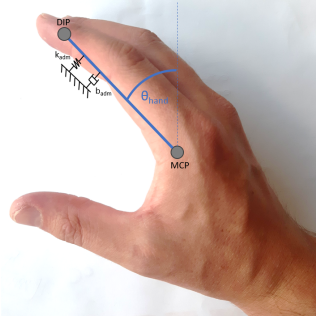


Fig. 5. Schematic of the Hand with the spring-damper system to generate a resistive force. The force will be generated by the device with the admittance control mode

In Figure 5 the above described idea can be seen. Depending on the user, the stiffness k_{adm} and b_{adm} can be changed by the therapist.

After starting the admittance control algorithm, the FSR sensors are calibrated and set to zero. This has to be done, because weight and gravitational forces of the users hand would disturb the measurement. The stiffness and damping coefficients are set by the user or the therapist.

The resistive force for the admittance control algorithm can therefore be calculated as follows.

$$M_{res} = k_{adm}\Theta_{hand} + b_{adm}\dot{\Theta}_{hand} \quad (4)$$

$$F_{res} = \frac{M_{res}}{l} \quad (5)$$

With length l , that is the distance between the rotational axis (in the center of the potentiometer in the device) and the force interaction point. To simplify that, the force interaction point is set to the middle of the top cover plate and l is fixed with the value $l = 0.17m$. The coefficient k_{adm} (unit $\frac{Nm}{deg}$) describes the stiffness of the system, b_{adm} (unit $\frac{Nm}{deg/s}$) is the damping coefficient. A change of the angular position Θ_{hand} and the angular velocity $\dot{\Theta}_{hand}$ results in a torque M_{res} against the movement

direction. The torque is used to calculate the virtual resistive force F_{res} , that the user has to overcome.

When the force F_{user} applied by the user is higher than F_{res} the motor is driven in the force direction. The motor speed is controlled depending on the difference of the forces ΔF .

$$\Delta F = F_{user} - F_{res} \quad (6)$$

The higher ΔF , the higher angular velocity can be achieved. The system works in both directions (push or pull force).

The process chart in Figure 6 shows the used algorithm.

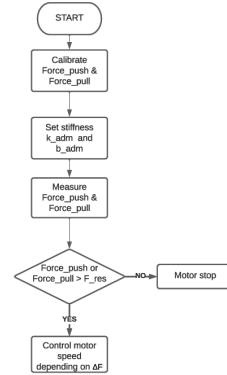


Fig. 6. Process chart of the admittance control algorithm

2) *Impedance control*: In order to enable an interaction between the rehabilitation device and the user an impedance control strategy is implemented. The goal is to achieve a haptic feedback when the patient is grabbing something in a virtual environment.

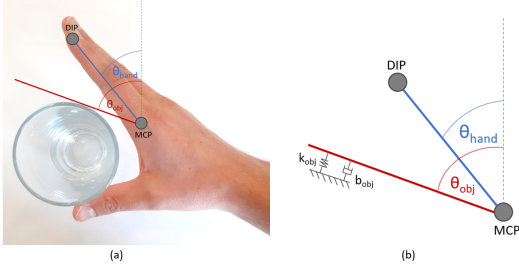


Fig. 7. (a) Schematic of the hand opening angle and the object contact angle. (b) Contact surface modeled with a spring-damper system

The contact surface (also called virtual wall) is modeled as a spring, and further with a combination of a spring and a damper. The spring constant k_{obj} depends on the stiffness of the simulated object. The contact force is calculated as follows.

$$\Theta_{diff} = \Theta_{obj} - \Theta_{hand} \quad (7)$$

$$M_{obj} = \begin{cases} 0 & \text{for } \Theta_{diff} > 0 \\ k_{obj}\Theta_{diff} + b_{obj}\dot{\Theta}_{diff} & \text{for } \Theta_{diff} \leq 0 \end{cases} \quad (8)$$

$$F_{obj} = \frac{M_{obj}}{l} \quad (9)$$

Where k_{obj} is the spring constant and b_{obj} the damping coefficient of the spring-damper system, as illustrated in Figure 7. When the user is moving it's hand towards the object, the resistance stays very low until the virtual object is touched. From this point, the force F_{obj} in opposite moving direction is generated and the user feels the resistance. The stiffness is chosen, depending on the simulated object.

To start the algorithm, the FSR sensors are calibrated (as mentioned in Section II-C1). After the initialisation, the main parameters of the object are defined (outer diameter, stiffness k_{obj}). With the outer diameter, the angle Θ_{obj} can be calculated. If the user is not touching the object, the motor is

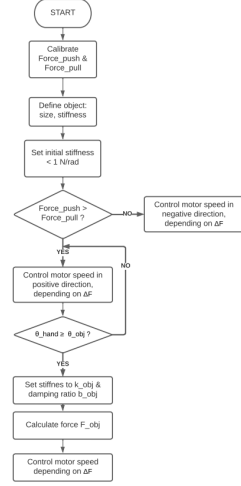


Fig. 8. Process chart of the impedance control algorithm

turned off so that the fingers can be moved with the minimum resistance. When the angle of the hand is the same or bigger than Θ_{obj} the resistive force is set (as can be seen in equation 8). Depending on ΔF , the motor speed is controlled in the direction of the force. The force difference for the impedance control is calculated as in equation 6 with F_{obj} instead of F_{res} .

D. Adaptive assistance - Assistance as needed

In this mode the motor is disabled, and the user can close or open its hand with a minimal resistive force. The resistance is mostly caused by the friction of the gears and the joints of the opening mechanism. A target position is set to a certain value and the user should try to open or close the hand in the direction of the target position. If the user is not able to move the fingers anymore, the motor is accelerating to a certain velocity and therefore helping the patient to reach the target position.

The software is measuring the actual angle every 3 seconds. If the angle at $t_1 = 0s$ and $t_2 = 3s$ is still the same (with a tolerance of $\pm 1^\circ$) the assistance is started and the motor is accelerating. The motor stops, when the target angle is reached.

By reaching the target angle, a new one is randomly generated within the working range of 5° and 45° . Thus, the therapy can be carried out continuously. Different parameters can be measured within this mode - f.e. the needed time to reach the target angles, the maximum velocity and the amount of assistance needed.

E. Adaptive assistance - User intention

This mode can use to mobilize patients with higher limitations but also advanced patients. The basic idea is that the user has to apply a certain force (push or pull) to move the device. Unlike the admittance control strategy, the device is moving all the way to the minimum or maximum angle if the force threshold is reached. The starting angle is fixed at $\Theta_{hand} = 25^\circ$, if the user is able to cross the force threshold by pushing on the door part (closing the hand), the device is moving to the minimum closing angle of $\Theta_{hand} = 5^\circ$. If the force is applied in the pulling direction, the device is opening to an angle of $\Theta_{hand} = 45^\circ$. The force threshold can be set $F_{tresh} = 2N$ for more severe patients and at $F_{tresh} = 30N$ for advanced training effects.

III. RESULTS

A. Admittance control

To evaluate the admittance control algorithm, the resistive force F_{res} , the force applied by the user F_{user} and the opening angle of the hand Θ_{hand} are recorded. The results are illustrated in Figure 9.

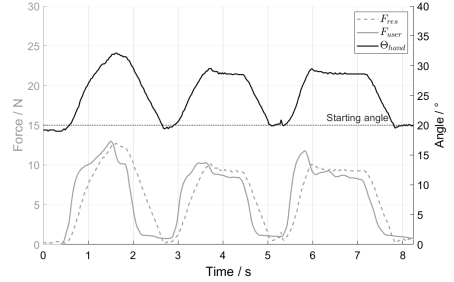


Fig. 9. Evaluation of the admittance control algorithm with a stiffness $k_{adm} = 1 \text{ Nm/deg}$

The angle is changing, when the force applied by the user is higher than the threshold force F_{res} . As example, this can be seen from second 0.5 to 1.5. Because the force applied by the user is higher than the resistive force, the device is moving from 20° to 32° . The bigger the force difference, the faster the angle is changing because the motor speed is higher. It can be seen, that the motor is blocking when the applied force F_{user} is lower than F_{res} . If the user is not able to hold the force, the motor is driving back to the starting position (dashed horizontal line).

B. Impedance control

For the evaluation of the impedance control algorithm three object with different spring stiffness constants are simulated.

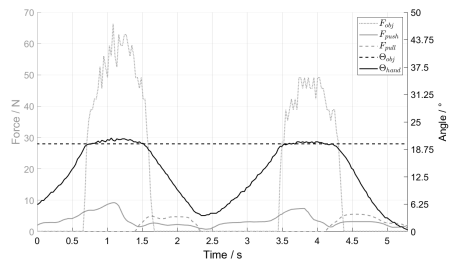


Fig. 10. Evaluation of the impedance control algorithm with a simulated hard object (stiffness $k_{obj} = 20 \text{ Nm/deg}$)

Figure 10 shows a hard object with with a stiffness of $k_{obj} = 20 \text{ Nm/deg}$. The simulated object should imitate materials like ceramic or glass.

It can be seen, that the resistive force F_{obj} is increasing rapidly, when the contact between the hand and the object is recognized. Due to that high step of the resistive force, the user is not able move the fingers above Θ_{obj} . This can be seen from $t_1 = 0.7 \text{ s}$ to $t_2 = 1.7 \text{ s}$ and $t_3 = 3.5 \text{ s}$ to $t_4 = 4.2 \text{ s}$. If the angle of the hand/fingers is lower than the object angle, the user is able to move the device freely in the two possible directions.

The lower the stiffness of the object, the more compression is possible. Figure 11 shows a very soft object that can be compressed with a lower force impact. It can be seen, that the user is able to compress the object up to 12° with a generated load of about 6.5 N .

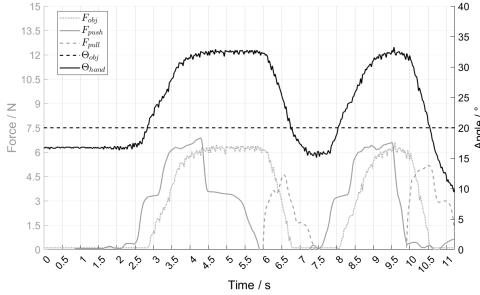


Fig. 11. Evaluation of the impedance control algorithm with a simulated soft object (stiffness $k_{obj} = 0.5 \text{ Nm/deg}$)

C. Assistance as needed mode

The resulting measurements of the assistance as needed mode can be seen in Figure 12. The user is trying to move the hand to the target position (dashed line). If this position could be reached, a new one is generated randomly between the 5° and 45° .

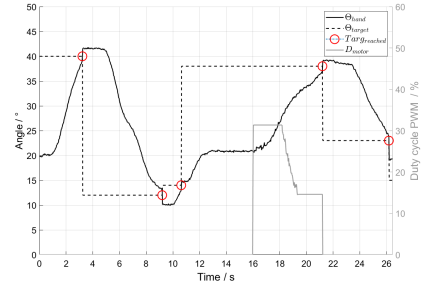


Fig. 12. Evaluation of the assistance as needed mode

This can be seen between $t = 13 \text{ s}$ and $t = 21 \text{ s}$. The user is not able to move the hand anymore (from $t = 13 \text{ s}$ to $t = 16 \text{ s}$ so that the position of the fingers does not change. After 3 seconds ($t = 16 \text{ s}$), the software recognises the stagnation and the motor is turned on to help the user. The assistance by the motor is turned off when the target angle is reached.

D. User intention mode

For the evaluation of the user intention mode, a force threshold F_{tresh} is set to 5 N .

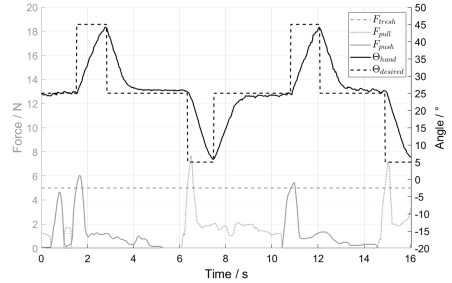


Fig. 13. Evaluation of the user intention mode with a force threshold of $F_{tresh} = 5 \text{ N}$

The opening angle of the device Θ_{hand} , the desired angle $\Theta_{desired}$ for the position control and the forces $F_{pull/push}$ are logged. The result can be

seen in Figure 13. If the user is able to cross the threshold by applying enough force, the device starts to move until the desired angle. Depending on the force direction, the desired angle is set to 5° (pulling force) or 45° (pushing force). This can be seen in the Figure above. At the time $t_1 = 1.8\text{ s}$, the force threshold is crossed as the user applying a pushing force of about $F_{push} = 6\text{ N}$. Due to that, the device is moving to an angle of $\Theta_{desired} = 45^\circ$.

E. Conclusion

In this thesis further developments were done for a basic version of an PHRD. A modular hardware was developed, to measure the gripping forces that are applied by the user.

With the developed system, the therapy can be performed interactively. The patient now has the possibility to control the device with his input.

With the admittance control method a training affect could be achieved as there is a resistance against the movement of the user. The resistance can be changed by adapting the stiffness coefficient in the software.

The impedance control shows the basic functions to simulate the gripping or touching of objects. With the implemented algorithm, the device responds when a virtual object with a defined stiffness is touched.

Due to the various modes, patients in different stages can be treated. The user intention mode can be used for patients with more severe impairments, as the patient specific force threshold has to be crossed only one time to make the complete movement.

When the patients have limitations of the ROM, the assistance as needed mode can be used. A target angle is generated automatically by the software, which the user has to reach. If the user is unable to move the fingers within a defined time, the device is assisting and moving in the direction of the target angle.

With the implemented control methods, a haptic feedback can be generated. This makes the therapy

more realistic, as movements of the daily life can be simulated.

ACKNOWLEDGMENT

I want to thank my supervisor Yeongmi Kim and my colleague Sebastian Seewald for their support during the master thesis.

REFERENCES

- [1] H. A. Wafa, C. D. A. Wolfe, E. Emmett, G. A. Roth, C. O. Johnson, and Y. Wang, "Burden of stroke in europe: Thirty-year projections of incidence, prevalence, deaths, and disability-adjusted life years," *Stroke*, vol. 51, no. 8, pp. 2418–2427, 2020.
- [2] V. L. Feigin, B. Norrving, and G. A. Mensah, "Global burden of stroke," *Circulation research*, vol. 120, no. 3, pp. 439–448, 2017.
- [3] S.-M. Lai, S. Studenski, P. W. Duncan, and S. Perera, "Persisting consequences of stroke measured by the stroke impact scale," *Stroke*, vol. 33, no. 7, pp. 1840–1844, 2002.
- [4] D. J. Clarke, L.-J. Burton, S. F. Tyson, H. Rodgers, A. Drummond, R. Palmer, A. Hoffman, M. Prescott, P. Tyrrell, L. Brkic, K. Grenfell, and A. Forster, "Why do stroke survivors not receive recommended amounts of active therapy? findings from the react study, a mixed-methods case-study evaluation in eight stroke units," *Clinical rehabilitation*, vol. 32, no. 8, pp. 1119–1132, 2018.
- [5] J. M. Veerbeek, A. C. Langbroek-Amersfoort, E. E. H. van Wegen, C. G. M. Meskers, and G. Kwakkel, "Effects of robot-assisted therapy for the upper limb after stroke," *Neurorehabilitation and neural repair*, vol. 31, no. 2, pp. 107–121, 2017.



Simon Deiser is a Master Student of Mechatronics at MCI Innsbruck/Austria. Previously finished the Bachelor in Mechatronics at MCI. Besides the studies, he is employed in the sales department of an industrial company.

Model Predictive Control of an Electrodynamic Shaker

Simon Eder and Davide Bagnara (supervisor)

Abstract—Vibration testing has become an essential part for the durability and reliability validation of machinery systems or structures. Sometimes, real occurring vibration profiles and properties are known from e.g. acceleration measurements collected during test runs. The here presented control strategy aims to directly use those profiles or any other sequence with certain constraints as reference excitation input. Hence, most realistic vibration tests can be performed.

By nature, electrodynamic shakers have a certain operation frequency and force range which define the constraints on the reference trajectory. Next to the constraint definition for set-point pre-filtering purposes, the development of a control algorithm with good trajectory tracing performance requires a detailed mathematical model of the system as well. However, the high computational effort accompanied by increasing model deepness poses a major challenge when it comes to running the control algorithm at the targeted control loop frequency of 20 kHz.

The system dynamics were derived from a lumped parameter model with *Lagrange's formalism*, where model parameter values come from various measurements. Performed frequency response measurements on the real shaker were used as a golden reference. Extending the standard excitation winding model proposed in literature by an additional resistor-inductor branch finally led to the desired model accuracy. Based on the dynamic model, a non-constrained model predictive control algorithm was programmed in MATLAB and subsequently translated into the Simulink environment. System simulations tests have shown very good trajectory tracking performance. Having a realistic dynamic model of the shaker showed its importance for simulations but also for the control algorithm itself. It allows to check beforehand if a trajectory can be tracked whilst complying certain shaker constraints or even automatically shape it to a realizable trajectory by applying the developed set-point pre-filter. Hence, the computational less demanding non-constrained MPC algorithm proved to be the right choice in terms of

control quality and computational feasibility.

This work has shown that the application field of MPC is not necessarily limited to slow processes. An optimal control technique was adopted to control a plant with a bandwidth in the range of several kilohertz with impressive reference tracking capabilities as simulations have demonstrated. If in future steps the proposed algorithm gets deployed on real hardware, using a transfer function model instead of the state space model for the predictions significantly reduces the computational effort. Moreover, full state observation becomes needless, the algorithm complexity reduces and the demands on the computation hardware diminish. This might flatten the way for MPC in the field of vibration testing and allow to excite the test structures most realistically.

Index Terms—model predictive control, modelling, electrodynamic shaker, set-point pre-filtering, simulation, trajectory tracking.

I SYSTEM MODELLING

I-A *The Plant's Physics - The Moving Coil Transducer Principle*

IN principle, an electrodynamic shaker can be seen as a moving coil transducer. Starting with this simplified and well studied system in e.g. [1], the transformation between electrical and mechanical energy and vice versa can be understood by applying fundamental physical concepts from electrodynamics. The hereby found coupling relations between the electrical and mechanical energies are the same as in the on hand electrodynamic shaker.

Fig. 1 depicts a moving coil transducer.

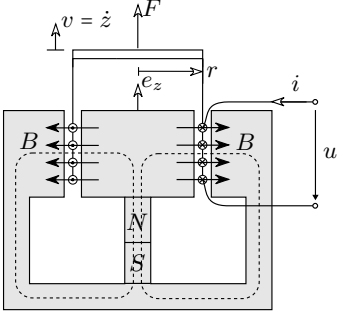


Fig. 1: Moving coil transducer [2]

The moving coil is placed in the air gap which, ideally, is a homogeneous and constant magnetic field \mathbf{B} , perpendicular to the moving direction e_x . The current i through the coil windings is driven from the excitation voltage u . Thus, applying the *Lorentz* force law which is defined as

$$\vec{F} = q(\mathbf{E} + \vec{v}_t \times \mathbf{B}) \quad (1)$$

results in a force acting on the armature winding. For the problem at hand the electric field \mathbf{E} can be neglected. The velocity \vec{v}_t of the charge q in the magnetic field \mathbf{B} can be split up in $\vec{v}_t = \vec{v} + \vec{v}_q$, where \vec{v} is due to the axial movement of the coil itself and \vec{v}_q is due to the charge movement along the ideally flat stacked coil wires (see Fig. 2).

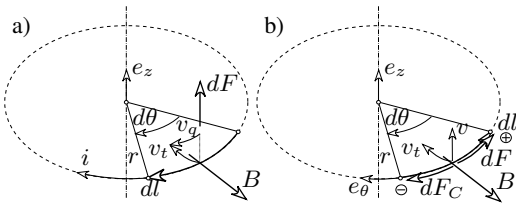


Fig. 2: *Lorentz* force on an infinitesimal conductor element: a) in axial direction, b) in circumferential direction [2]

According to that split in Fig. 2(a) we obtain the force in axial direction for an infinitesimal

conductor element as follows:

$$d\vec{F} = dF\vec{e}_z = dq\vec{v}_q \times \mathbf{B} = dq \frac{d\vec{l}}{dt} \times \mathbf{B} = id\vec{l} \times \mathbf{B} \quad (2)$$

given that $d\vec{l}$ is perpendicular to \mathbf{B} ,

$$d\vec{F} = ird\theta B\vec{e}_z. \quad (3)$$

To obtain the magnitude of resulting force in axial direction, Eqn. 3 is integrated over the length of the coil $L = 2\pi rn$, where n counts the winding turns. The force in Fig. 1 represents the external force to balance the magnetic force. Hence, we have to change the sign to obtain

$$F = -2r\pi nBi. \quad (4)$$

As a consequence of this force and the resulting axial motion of the conductor in the \mathbf{B} -field, the force in the direction of the conductor (Fig. 2(b)) is:

$$d\vec{F} = dF\vec{e}_\theta = -dq(\vec{v} \times \mathbf{B}) \quad (5)$$

given that \vec{v} is perpendicular to \mathbf{B} ,

$$d\vec{F} = -dqvB\vec{e}_\theta. \quad (6)$$

Since the *Coulomb* force acts in opposite direction ($d\vec{F}_C = -d\vec{F}$), an electric field builds up as follows:

$$\mathbf{E} = \frac{d\vec{F}_C}{dq} = -\frac{d\vec{F}}{dq} = vB\vec{e}_\theta. \quad (7)$$

The induced voltage of an infinitesimal conductor element in the direction of the current reads as

$$du = \mathbf{E} \cdot d\vec{l} = vBr d\theta. \quad (8)$$

and integrating over the length of the coil, we finally obtain the so called back-emf as

$$u_{\text{back emf}} = 2r\pi nBv. \quad (9)$$

With the substitution

$$M = 2r\pi nB, \quad (10)$$

the constitutive relations in Eqn. 4 and Eqn. 10 for the moving coil can be expressed in a better readable form as

$$F = -Mi \quad (11)$$

and

$$u_{\text{back emf}} = Mv. \quad (12)$$

The transducer constant M appearing in Eqn. 11 and Eqn. 12 are the same. This is because the moving coil transducer is considered to be ideal.

I-B Extended Moving Coil Impedance Model

As proposed in [1] and [3], the moving coil impedance is usually modeled as a serial R-L branch. However, measurements have shown that this simple model is not an adequate representation of the real shaker's moving coil impedance over the frequency range of interest. Thus, including a parallel R-L branch in series to the serial R-L branch (Fig. 3) allows to display the low frequency range much better by just adding a single current path.

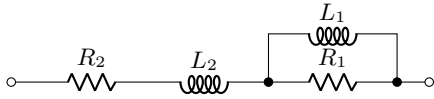


Fig. 3: Moving coil equivalent lumped element model

Applying the MATLAB "lsqcurvefit" function on the impedance function

$$Z(j\omega) = R_2 + \frac{\omega^2 L_1^2 R_1}{R_1^2 + \omega^2 L_1^2} + j \left(\omega L_2 + \frac{\omega L_1 R_1^2}{R_1^2 + \omega^2 L_1^2} \right) \quad (13)$$

for the circuit in Fig. 3, gives the values $R_1 = 1.775 \Omega$, $L_1 = 3.853 \text{ mH}$, $R_2 = 1.680 \Omega$ and $L_2 = 0.273 \text{ mH}$. The impedance comparison of this model with the real moving coil is shown in Fig. 4.

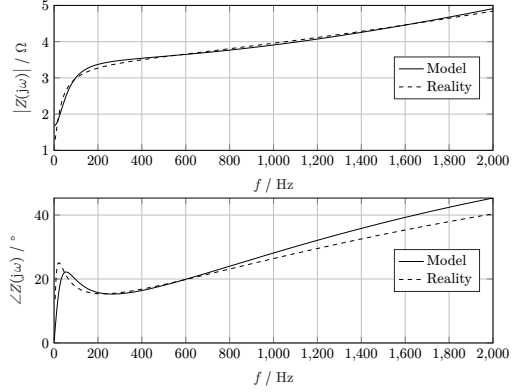


Fig. 4: Magnitude and phase comparison between R-L-R-L model and real excitation coil impedance

Fig. 5 shows the lumped element model on which the dynamic equations were derived using *Lagrange's formalism*. Eqn. 14 is the result in state space form, including also the power amplifier dynamics, modelled as a PT1 system. The derivation process is shown in detail in [4].

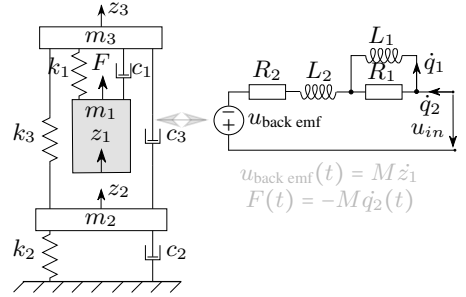


Fig. 5: Lumped element model of the at hand electrodynamic shaker

$$\begin{bmatrix} \dot{z}_1 \\ \dot{z}_2 \\ \dot{z}_3 \\ \dot{q}_1 \\ \dot{q}_2 \end{bmatrix} = \begin{bmatrix} -1/r & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & -R_1/L_2 & R_1/L_1 & 0 & 0 \\ 1/L_2 & 0 & 0 & 0 & -R_1/L_2 & -(R_1 + R_2)/L_2 & -M/L_2 & M/L_2 & 0 \\ 0 & -k_1/m_1 & 0 & k_1/m_1 & 0 & -M/m_1 & -c_1/m_1 & 0 & c_1/m_1 \\ 0 & 0 & -(k_2 + k_3)/m_2 & k_3/m_2 & 0 & -M/m_2 & 0 & -(c_2 + c_3)/m_2 & c_3/m_2 \\ 0 & k_1/m_3 & k_3/m_3 & -(c_1 + c_3)/m_3 & 0 & 0 & c_1/m_3 & c_3/m_3 & -(c_1 + c_3)/m_3 \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \\ z_3 \\ q_1 \\ q_2 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \begin{bmatrix} u_{\text{back emf}} \\ F(t) \\ u_{\text{in}} \end{bmatrix} \quad (14)$$

In in short, Eqn. 14 writes

$$\dot{\tilde{\mathbf{x}}} = \mathbf{A}\tilde{\mathbf{x}} + \tilde{\mathbf{b}}u_c. \quad (15)$$

I-C Model Validation

Tab. I summarizes all the parameters used in Eqn. 14. The values of the mechanical and electrical parameters were obtained by measurements. The electromechanical coupling coefficient was calculated based on data provided by the shaker manufacturer and verified by means of a magnetic transient simulation using Ansys Maxwell (see [4]).

TABLE I: Summary table of the shaker model parameters

Parameter	Symbol	Value
Power amplifier gain	G	29.5
Power amplifier time constant	τ	30 μs
Moving coil resistance 1	R_1	1.775 Ω
Moving coil resistance 2	R_2	1.680 Ω
Moving coil inductance 1	L_1	3.853 mH
Moving coil inductance 2	L_2	0.273 mH
Electromechanical coupling coefficient	M	-51 V s m^{-1}
Moving coil stiffness	k_1	315.150 MN m^{-1}
Moving coil damping	c_1	120 kg s^{-1}
Body stiffness	k_2	160.017 kN m^{-1}
Body damping	c_2	681.492 kg s^{-1}
Armature stiffness	k_3	24.407 kN m^{-1}
Armature damping	c_3	37.76 kg s^{-1}

As a quality measure, frequency response functions (FRF's) of the model and the real shaker are compared. The FRF used for comparison was obtained from the control input signal u_c and the armature acceleration \ddot{z}_3 . The algorithms used for the estimation of the FRF's are based on [5]. The comparison is finally shown in Fig. 6 and Fig. 7.

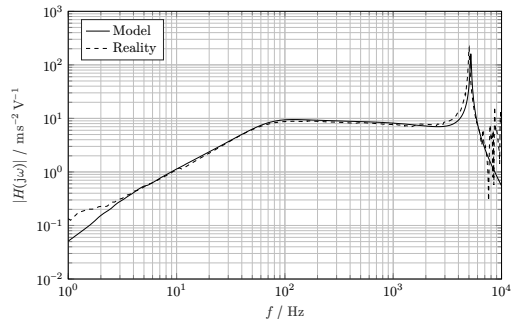


Fig. 6: Magnitude response comparison between model and real shaker

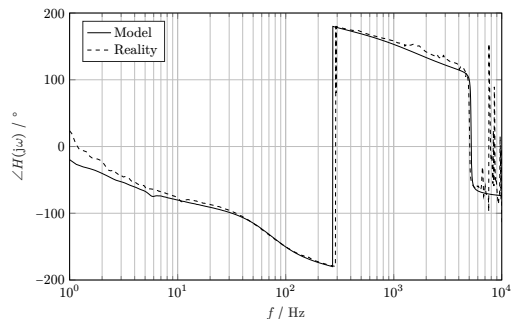


Fig. 7: Phase response comparison between model and real shaker

A very good reality matching model can be observed over a wide frequency range. Up to 2 kHz which poses the upper frequency bound for the input signal, reliable predictions can be made for the MPC algorithm.

II STATE SPACE FORMULATION OF UNCONSTRAINED MPC

In the following, the control law for the unconstrained MPC is derived for linear time invariant systems given in state space form. Despite the fact that we only have to deal with a SISO system, the control law is derived for the general case and thus also applicable on MIMO systems. Tab. II shows the

explanation of some important parameters used in this section.

TABLE II: Parameter description list for this section

Parameter	Description
n	number of system states
m	number of controlled states
H_p	higher value of prediction horizon
H_w	lower value of prediction horizon
H_c	control horizon

Note that the continuous time model (Eqn. 14) must first be converted into a discrete time model. This is done by the MATLAB "c2d" command. The sampling time t_s required to pass to this function was chosen 1/20 kHz; hence, being one order of magnitude beyond the fastest dynamic aimed to control. Now let $\hat{\mathbf{x}}(k+i|k)$ ($1 \leq i \leq H_p$) be the predicted state vector of the system i time steps from the current time step k . Then, according to [6], H_p predictions steps can be computed by iterating the discrete model as follows:

$$\begin{bmatrix} \hat{\mathbf{x}}(k+1|k) \\ \hat{\mathbf{x}}(k+2|k) \\ \hat{\mathbf{x}}(k+3|k) \\ \vdots \\ \hat{\mathbf{x}}(k+H_w|k) \\ \hat{\mathbf{x}}(k+H_w+1|k) \\ \vdots \\ \hat{\mathbf{x}}(k+H_c|k) \end{bmatrix} = \begin{bmatrix} \mathbf{A} \\ \mathbf{A}^2 \\ \mathbf{A}^3 \\ \vdots \\ \mathbf{A}^{H_w} \\ \mathbf{A}^{H_w+1} \\ \vdots \\ \mathbf{A}^{H_c} \end{bmatrix} \hat{\mathbf{x}}(k) + \begin{bmatrix} (\mathbf{I}-\mathbf{A})\mathbf{B} \\ (\mathbf{I}-\mathbf{A}+\mathbf{A}^2)\mathbf{B} \\ \vdots \\ \sum_{i=0}^{H_w-1} \mathbf{A}^i \mathbf{B} \\ \vdots \\ \sum_{i=0}^{H_c-H_w-1} \mathbf{A}^i \mathbf{B} \end{bmatrix} \bar{\mathbf{u}}(k-1) + \begin{bmatrix} \mathbf{B} & 0 & \dots & 0 \\ (\mathbf{I}-\mathbf{A})\mathbf{B} & \mathbf{B} & \dots & 0 \\ (\mathbf{I}-\mathbf{A}+\mathbf{A}^2)\mathbf{B} & (\mathbf{I}-\mathbf{A})\mathbf{B} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=0}^{H_w-2} \mathbf{A}^i \mathbf{B} & \sum_{i=0}^{H_w-1} \mathbf{A}^i \mathbf{B} & \dots & \mathbf{B} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=0}^{H_w-1} \mathbf{A}^i \mathbf{B} & \sum_{i=0}^{H_w-1} \mathbf{A}^i \mathbf{B} & \dots & (\mathbf{I}-\mathbf{A})\mathbf{B} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=0}^{H_c-2} \mathbf{A}^i \mathbf{B} & \sum_{i=0}^{H_c-2} \mathbf{A}^i \mathbf{B} & \dots & \sum_{i=0}^{H_c-H_c} \mathbf{A}^i \mathbf{B} \end{bmatrix} \begin{bmatrix} \Delta \hat{\mathbf{u}}(k|k) \\ \Delta \hat{\mathbf{u}}(k+1|k) \\ \Delta \hat{\mathbf{u}}(k+2|k) \\ \vdots \\ \Delta \hat{\mathbf{u}}(k+H_c-1|k) \end{bmatrix}$$

with $\Delta \hat{\mathbf{u}}(k+i|k)$ ($0 \leq i \leq H_c-1$) being the applied input increments. And in a better to read form we write this equation as

$$\hat{\mathbf{X}}(k) = \mathbf{F}\hat{\mathbf{x}}(k) + \mathbf{G}\bar{\mathbf{u}}(k-1) + \mathbf{H}\Delta \hat{\mathbf{U}}(k). \quad (16)$$

Defining the tracking error (Eqn. 17) would allow to calculate the deviation from the reference, occurring, if the system was in free response, meaning no input would be applied; i.e. $\Delta \hat{\mathbf{U}}(k) = \mathbf{0}$.

$$\bar{\mathbf{e}}(k) = \bar{\mathbf{X}}_{\text{ref}}(k) - \mathbf{F}\hat{\mathbf{x}}(k) - \mathbf{G}\bar{\mathbf{u}}(k-1) \quad (17)$$

The goal, however, is to find the inputs $\Delta \hat{\mathbf{U}}(k)$ that minimize the cost function

$$\bar{\mathbf{J}}(k) = \left| \hat{\mathbf{X}}(k) - \bar{\mathbf{X}}_{\text{ref}}(k) \right|^2 \mathbf{Q} + \left| \Delta \hat{\mathbf{U}}(k) \right|^2 \mathbf{R}. \quad (18)$$

By further substitutions with Eqn. 16 and Eqn. 17 we get:

$$\bar{\mathbf{J}}(k) = \left| \mathbf{H}\Delta \hat{\mathbf{U}}(k) - \bar{\mathbf{e}}(k) \right|^2 \mathbf{Q} + \left| \Delta \hat{\mathbf{U}}(k) \right|^2 \mathbf{R} \quad (19)$$

\mathbf{Q} and \mathbf{R} are the so called weighting matrices. \mathbf{Q} is used to penalize for the tracking error, whereas \mathbf{R} penalizes the control inputs. Beside assigning different penalizing factors for each state's error, it is also possible to prioritize the minimization of the state's error of each prediction step from (H_w to H_p). The same also applies for the control input increments from step 0 up to H_c . Hence, in addition to the control and prediction horizons, these two matrices pose two more parameters for the controller tuning. In practice, however, the diagonal entries in the \mathbf{Q} matrix are often just set to one, and the diagonal entries in the \mathbf{R} matrix are chosen very small or even zero in order to prioritize the reduction of the state's error over the whole prediction horizon equally.

Now there are different ways of finding the set of control inputs $\Delta \hat{\mathbf{U}}(k)$ which minimize the cost function Eqn. 19. In [6] it is proposed to find the solution by formulating it as a least squares problem. The reason is that the \mathbf{H} matrix is sometimes ill conditioned, which can lead to numerical problems when finding the optimal solution.

Therefore, we start with the least squares formulation by finding the matrices \mathbf{S}_Q and \mathbf{S}_R by extracting the square root of \mathbf{Q} and \mathbf{R} , which we can do since $\mathbf{Q} \geq 0$ and $\mathbf{R} \geq 0$:

$$\mathbf{S}_Q^T \mathbf{S}_Q = \mathbf{Q} \quad \mathbf{S}_R^T \mathbf{S}_R = \mathbf{R} \quad (20)$$

If \mathbf{Q} and \mathbf{R} are diagonal, just taking the square root of each diagonal element gives the result. For the case where they are not diagonal, [6] proposes the "Cholesky" algorithm to obtain the square-roots of

positive definite matrices or singular value decomposition (svd) for semi-definite matrices.

Now we define the vector

$$\begin{bmatrix} \mathbf{S}_Q \left(\mathbf{H} \Delta \hat{\mathbf{U}}(k) - \bar{\boldsymbol{\epsilon}}(k) \right) \\ \mathbf{S}_R \Delta \hat{\mathbf{U}}(k) \end{bmatrix},$$

which basically consists of Eqn. 19 first term's square root in the first row and its second term's square root in the second row. Subsequently it is shown that the squared length of this vector ($\|\cdot\|^2$) i.e. the sum of squares, is the cost function $\bar{\mathbf{J}}(k)$ itself.

$$\begin{aligned} & \left\| \begin{bmatrix} \mathbf{S}_Q \left(\mathbf{H} \Delta \hat{\mathbf{U}}(k) - \bar{\boldsymbol{\epsilon}}(k) \right) \\ \mathbf{S}_R \Delta \hat{\mathbf{U}}(k) \end{bmatrix} \right\|^2 = \\ & = \left\| \begin{bmatrix} \mathbf{S}_Q \left(\bar{\mathbf{X}}(k) - \bar{\mathbf{X}}_{\text{ref}}(k) \right) \\ \mathbf{S}_R \Delta \hat{\mathbf{U}}(k) \end{bmatrix} \right\|^2 = \\ & = \left(\bar{\mathbf{X}}(k) - \bar{\mathbf{X}}_{\text{ref}}(k) \right)^\top \mathbf{S}_Q^\top \mathbf{S}_Q \left(\bar{\mathbf{X}}(k) - \bar{\mathbf{X}}_{\text{ref}}(k) \right) \\ & \quad + \Delta \hat{\mathbf{U}}(k)^\top \mathbf{S}_R^\top \mathbf{S}_R \Delta \hat{\mathbf{U}}(k) = \\ & = \left| \bar{\mathbf{X}}(k) - \bar{\mathbf{X}}_{\text{ref}}(k) \right|^2 \mathbf{Q} + \left| \Delta \hat{\mathbf{U}}(k) \right|^2 \mathbf{R} = \bar{\mathbf{J}}(k) \end{aligned} \quad (21)$$

Thus, the set $\Delta \hat{\mathbf{U}}(k)$ which minimizes the cost ($\bar{\mathbf{J}}(k) \stackrel{\dagger}{=} \bar{\mathbf{0}}$) is the least-squares solution of this problem and simultaneously the optimal control input $\Delta \hat{\mathbf{U}}(k)_{\text{opt}}$. Hence, we set

$$\begin{bmatrix} \mathbf{S}_Q \left(\mathbf{H} \Delta \hat{\mathbf{U}}(k) - \bar{\boldsymbol{\epsilon}}(k) \right) \\ \mathbf{S}_R \Delta \hat{\mathbf{U}}(k) \end{bmatrix} = \bar{\mathbf{0}}, \quad (22)$$

which is equivalent to

$$\begin{bmatrix} \mathbf{S}_Q \mathbf{H} \\ \mathbf{S}_R \end{bmatrix} \Delta \hat{\mathbf{U}}(k) = \begin{bmatrix} \mathbf{S}_Q \bar{\boldsymbol{\epsilon}}(k) \\ \bar{\mathbf{0}} \end{bmatrix}. \quad (23)$$

This equation has the form $\mathbf{A}\bar{\boldsymbol{x}} = \bar{\boldsymbol{b}}$ and can be solved in MATLAB with the command $\bar{\boldsymbol{x}} = \mathbf{A} \setminus \bar{\boldsymbol{b}}$. Using

this command is advantageous because it computes the solution regardless badly scaled or nearly singular \mathbf{A} matrices (see MATLAB documentation). Thus, the optimal control increments using MATLAB notation are

$$\Delta \hat{\mathbf{U}}(k)_{\text{opt}} = \begin{bmatrix} \mathbf{S}_Q \mathbf{H} \\ \mathbf{S}_R \end{bmatrix} \setminus \begin{bmatrix} \mathbf{S}_Q \bar{\boldsymbol{\epsilon}}(k) \\ \bar{\mathbf{0}} \end{bmatrix}. \quad (24)$$

Since we aim to implement the unconstrained MPC algorithm according to Fig. 8, the control matrix \mathbf{K}_{MPC} is the last missing object required. Again using MATLAB notation and calculating the matrix \mathbf{K} , which appears to be the right side of Eqn. 24 but without $\bar{\boldsymbol{\epsilon}}$

$$\mathbf{K} = \begin{bmatrix} \mathbf{S}_Q \mathbf{H} \\ \mathbf{S}_R \end{bmatrix} \setminus \begin{bmatrix} \mathbf{S}_Q \\ \bar{\mathbf{0}} \end{bmatrix} \quad (25)$$

and extracting only its first n rows finally gives (in MATLAB notation)

$$\mathbf{K}_{\text{MPC}} = \mathbf{K}(1:n, :). \quad (26)$$

The reason for using only the first n rows derives from the fact that at each time step k , one control move is sent to the plant. Then, at the next time step, the states are measured again, new predictions are made and the error $\bar{\boldsymbol{\epsilon}}(k)$ is calculated accordingly; multiplying the error with the gain matrix \mathbf{K}_{MPC} again gives the next, optimal control move increment. This procedure always repeats.

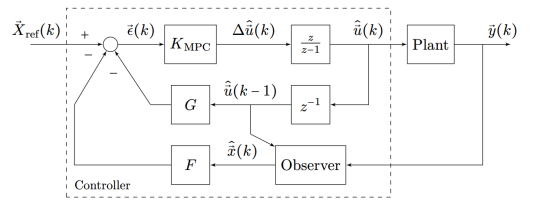


Fig. 8: Unconstrained MPC structure

Up to this point we have assumed that we have access to all systems states through measurements. Since this is not the case in reality, a state observer is needed in order to get a full state estimate (see

Sec. II-A). Fortunately, this does not require to change the derived control law. Replacing the full state measurement $\tilde{\mathbf{x}}(k)$ with the full state estimate $\hat{\tilde{\mathbf{x}}}(k)$ calculated by the observer already fixes this issue (see Fig. 8).

II-A Reduced Order State Observer Design

For reasons of computational efficiency, a reduced order state observer according to [7] is used. Therefore, the state vector $\tilde{\mathbf{x}}$ is partitioned in the following two parts: $\tilde{\mathbf{x}}_a$, the directly measured portion and $\tilde{\mathbf{x}}_b$, the unmeasured portion to be estimated. This results in the following system description:

$$\begin{bmatrix} \tilde{\mathbf{x}}_a(k+1) \\ \tilde{\mathbf{x}}_b(k+1) \end{bmatrix} = \begin{bmatrix} \mathbf{A}_{aa} & \mathbf{A}_{ab} \\ \mathbf{A}_{ba} & \mathbf{A}_{bb} \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{x}}_a(k) \\ \tilde{\mathbf{x}}_b(k) \end{bmatrix} + \begin{bmatrix} \tilde{\mathbf{b}}_a \\ \tilde{\mathbf{b}}_b \end{bmatrix} u_c(k). \quad (27)$$

$$\tilde{\mathbf{y}}(k) = \begin{bmatrix} \mathbf{I}_{3 \times 3} & \mathbf{0}_{3 \times 6} \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{x}}_a(k) \\ \tilde{\mathbf{x}}_b(k) \end{bmatrix} \quad (28)$$

Hence, according to [7], the reduced order observer equation writes:

$$\begin{aligned} \hat{\tilde{\mathbf{x}}}_b(k+1) &= \mathbf{A}_{bb} \hat{\tilde{\mathbf{x}}}_b(k) + \mathbf{A}_{ba} \tilde{\mathbf{x}}_a(k) + \tilde{\mathbf{b}}_b u_c(k) \\ &+ \mathbf{L}_r (\tilde{\mathbf{x}}_a(k+1) - \mathbf{A}_{aa} \tilde{\mathbf{x}}_a(k) \\ &- \tilde{\mathbf{b}}_a u_c(k) - \mathbf{A}_{ab} \hat{\tilde{\mathbf{x}}}_b(k)) \end{aligned} \quad (29)$$

The reduced observer gain matrix \mathbf{L}_r can be obtained the same way as for a standard full state observer. Thus, calculating the eigenvalues of the matrix \mathbf{A}_{bb} in MATLAB by

```
eig(A_bb)
```

gives the reduced system poles shown in Tab. III, which also contains the chosen poles of the reduced order observer.

TABLE III: Reduced system and observer poles in continuous and discrete time domain

System poles (CT)	Des. obs. poles (CT)	Des. obs. poles (DT)
$-154.719 + 28\,506.793j$	-928.313	0.955
$-154.719 - 28\,506.793j$	-928.313	0.954
$-2.989 + 39.040j$	-29.894	0.998
$-2.989 - 39.040j$	-29.894	0.998
0	$-1 \cdot 10^{-6}$	0.999
-460.680	-506.748	0.975

Consequently, the reduced observer gain matrix \mathbf{L}_r obtained by the MATLAB "place" command in the following way:

```
L_r = place(A_bb', A_ab', discrete desired reduced observer poles)'
```

Note that the matrices \mathbf{A}_{bb} and \mathbf{A}_{ab} in the above function call are in discretized form.

III SIMULATION RESULTS

Fig. 9, 10 and 11 show the results of loaded case simulations for three different reference trajectories, where a mass of 0.5 kg is placed on the shaker table. This additional mass can be seen as an external disturbance because it is neither included in the prediction model of the MPC algorithm nor in the state observer matrices. Thus, the controller's ability of compensating for such disturbances can be observed.

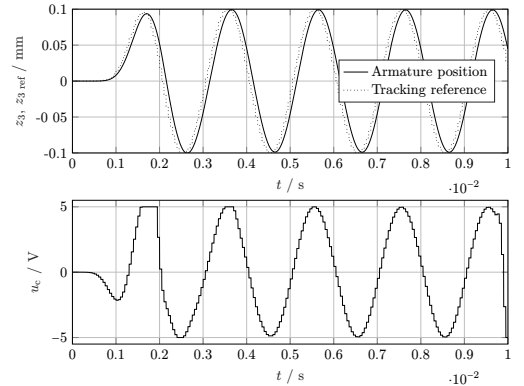


Fig. 9: 500 Hz sinusoidal tracking and control input

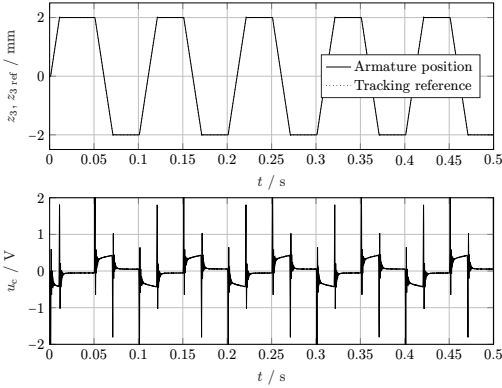


Fig. 10: Ramp tracking and control input

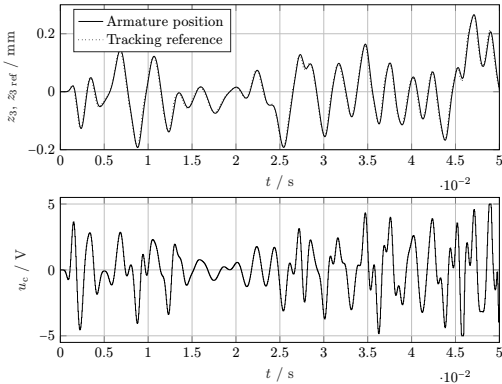


Fig. 11: Random tracking and control input

IV CONCLUSION

A novel representation of the excitation coil winding in form of an R-L-R-L branch has significantly reduced the error in the frequency responses between the model and the golden reference. In the frequency range of interest (1 Hz to 2 kHz) the model is very accurate. The decision to implement an MPC algorithm with unconstrained optimization was based on considerations about computation effort. Hence, the derivation of the control law was clearly elaborated

and explained. Furthermore, a computationally efficient state observer, namely, a reduced order state observer was implemented. Closed loop simulations where all single components (model, MPC, set-point pre-filter and state observer) must play together were performed in the Simulink environment. Simulations with different reference trajectories and external disturbances have demonstrated the high control quality of the MPC algorithm.

REFERENCES

- [1] A. Preumont, *Mechatronics: Dynamics of electromechanical and piezoelectric systems*, ser. Solid mechanics and its applications. Dordrecht: Springer, 2006, vol. 136.
- [2] Franz-Josef Falkner, "Modeling of a vibration test system," March 2022.
- [3] J. Martino and K. Harri, "Two-port modeling and simulation of an electrodynamic shaker for virtual shaker testing applications," *Journal of Sound and Vibration*, vol. 460, p. 114835, 2019.
- [4] Simon Eder, "Model Predictive Control of an Electrodynamic Shaker," Master Thesis, MCI - Management Center Innsbruck, Innsbruck, 2022.
- [5] H Vold, J Crowley, GT Rocklin, "New ways of estimating frequency response functions," vol. 18, pp. 34–38, 1984.
- [6] J. M. Maciejowski, *Predictive control: With constraints*, 1st ed. Harlow and Munich: Prentice Hall, 2002.
- [7] G. F. Franklin, J. D. Powell, and M. L. Workman, *Digital control of dynamic systems*, 3rd ed., ser. World student series. Menlo Park, Calif.: Addison-Wesley, 1998.

Simon Eder is a graduate mechatronics student at MCI Innsbruck/Austria.



=2

Produktionsoptimierung von Defibrillationselektroden mithilfe einer roboterunterstützten Teilautomatisierung

Daniel Federer, Benjamin Massow (Betreuer) und Armin Lechner (Betreuer)

Kurzfassung—Produzierende Unternehmen in Zentraleuropa stehen vermehrt vor dem Problem, Kundennachfragen nicht decken zu können. Ein Grund dafür ist der Mangel an Fachpersonal. Um dennoch konkurrenzfähig und rentabel produzieren zu können, werden Automatisierungslösungen benötigt. Produktionen werden dadurch skalierbarer. Zudem werden Mitarbeiter:innen für höherwertige Aufgaben frei. Durch den rasanten Anstieg der Nachfrage an Defibrillationselektroden sind die entsprechenden Produktionen überlastet. Ziel dieser Arbeit ist es, deren Produktion zu optimieren. Dazu wird mithilfe einer Wertstromanalyse der aktuelle Herstellprozess des Medizinprodukts aufgenommen und analysiert. Basierend darauf wird mit der Methodik des Wertstromdesigns ein verbesserter Produktionsablauf erarbeitet. Um den zu erreichen, bedarf es der schrittweisen Implementierung mehrerer verschiedener Verbesserungsprojekte. Eines davon ist die Entwicklung einer auf einem Robotersystem basierenden Produktionsmaschine. Die gestellten Anforderungen an die Maschine werden größtenteils bereits im Prototypenstatus erfüllt. Andere können durch in der Validierung erarbeitete Verbesserungsmaßnahmen in der Umsetzung der Produktionsmaschine erreicht werden. Daher kann die Machbarkeit einer auf dem Prototyp basierenden Produktionsmaschine bestätigt werden.

Schlagwörter—Robotik, Teilautomatisierung, Produktionsoptimierung

I. EINLEITUNG

DIE Produktion von Defibrillationselektroden stößt aufgrund der in den letzten Jahren stark

D. Federer studiert am Studiengang Mechatronik, MCI, Innsbruck, Österreich

gestiegenen Produktnachfrage an ihre Kapazitätsgrenzen. Kundenanfragen werden teilweise abgelehnt, da die Produktion die Nachfrage nicht decken kann. Aufgrund des hohen Anteils an manuell durchzuführenden Arbeiten ist die Skalierung der Produktion herausfordernd. Um die Produktionsmengen in solchen personalintensiven Produktionen zu erhöhen, werden neue Fachkräfte benötigt, die nur schwer zu finden sind. Daher sollen Automatisierungslösungen entwickelt werden. Dadurch wird die Produktion nicht nur skalierbarer, sondern auch das Personal kann entlastet werden.

In einem ersten Schritt wird der entsprechende Produktionsbereich analysiert. Defizite in der Produktion und Verbesserungspotentiale werden methodisch identifiziert, um Umsetzungsprojekte zur Optimierung der Produktion definieren zu können. Eines dieser Projekte umfasst die Entwicklung einer Produktionsmaschine, die den Einsatz von Robotersystemen vorsieht. In dieser Arbeit ist die Machbarkeit dieser Maschine zu prüfen. Dazu wird ein Prototyp konzipiert, aufgebaut und anhand definierter Randbedingungen getestet.

II. METHODIK

Eine Produktionsoptimierung kann insbesondere dann erreicht werden, wenn es gelingt, die aktuelle Situation einer Produktion darzustellen, den Ablauf und die Gewohnheiten systematisch in Frage zu stellen und zielorientiert weiterzuentwickeln [1]. Die Wertstrommethode ist dahingehend ein Werkzeug [2]. Mit ihr werden Prozessabläufe dargestellt und

ein Überblick der aktuellen Situation sowie des angestrebten Soll-Zustands einer Produktion geschaffen [1]. Der erarbeitete Soll-Zustand wird als Wertstromdesign bezeichnet, wobei zuvor die Ist-Situation mithilfe einer sogenannten Wertstromanalyse zu erfassen ist.

A. Wertstromanalyse

Mit der Wertstromanalyse wird der gesamte Wertstrom eines Produkts bzw. einer Produktgruppe von der Auslieferung an den Kunden bis zurück zur Anlieferung der Rohmaterialien erfasst [3], [4]. Der Wertstrom beschreibt dabei den Prozess der Wertschöpfung und umfasst alle wertschöpfenden und nicht wertschöpfenden Tätigkeiten, die zur Herstellung eines Produkts notwendig sind [1].

Die Durchführung einer Wertstromanalyse erfolgt in vier Schritten [1]:

1) *Produktfamilienbildung*: Zunächst ist das zu betrachtende Produkt bzw. falls mehrere Varianten existieren, die Produktgruppe zu definieren [5]. Ein einfaches und sicheres Verfahren dafür basiert auf der sogenannten Produktfamilien-Matrix, in welcher alle Produkte und die von ihnen durchlaufenden Prozessschritte tabellarisch erfasst werden. Anschließend können Produkte mit ähnlichen oder gleichen Prozessschritten gruppiert werden und bilden eine Produktfamilie.

2) *Kundenbedarfsanalyse*: Im zweiten Schritt der Wertstromanalyse wird der Kundentakt KT ermittelt. Dieser berechnet sich zu

$$KT = \frac{PT \cdot AZ}{KB} \quad (1)$$

mit den Produktionstagen PT , der täglichen effektiven Arbeitszeit AZ und der Absatzmenge des zurückliegenden Geschäftsjahres, dem sogenannten Kundenbedarf KB . Der Kundentakt gibt vor in welcher Zeit ein Produkt fertiggestellt werden muss, um die Nachfrage des Kunden erfüllen zu können [6].

3) *Wertstromaufnahme*: Nachdem das Produkt sowie der Kunde und dessen Bedarf bekannt sind, kann mit der Aufnahme des Wertstroms begonnen

werden. Dazu werden alle Produktionsprozesse sowie der damit verbundene Materialfluss im Zuge eines Rundgangs durch die Produktion skizziert. Hierbei ist aus Sicht des Kunden vorzugehen, d.h. der Rundgang beginnt beim Versand der Produkte und arbeitet sich rückwärts durch den Wertstrom bis zur Anlieferung der Rohmaterialien.

4) *Verbesserungspotentiale*: Ist die Wertstromaufnahme fertiggestellt, wird sie analysiert. Dabei werden die einzelnen Bearbeitungszeiten jedes Prozesses mit dem Kundentakt verglichen. Zu langsame Prozesse bzw. Prozesse oder Abläufe, die zu optimieren sind, werden anschließend im Wertstrom markiert.

B. Wertstromdesign

Sobald der Ist-Zustand in der Produktion erfasst ist und die Verbesserungspotentiale aufgezeigt sind, kann mit der Ausarbeitung des Soll-Zustands begonnen werden [7]. In [1] werden durch Erlach acht Gestaltungsrichtlinien definiert, die bei der Optimierung der Produktion unterstützen. Ziel des Wertstromdesigns ist es, eine schlanke und effiziente Produktion zu entwickeln, in der jegliche Art von Verschwendung (Muda) eliminiert wird [8].

C. Entwicklung mechatronischer Systeme

Um sich dem im Wertstromdesign erarbeiteten Soll-Zustand schrittweise anzunähern, sind Verbesserungsprojekte umzusetzen. Bei produzierenden Unternehmen erfordern solche Projekte meist die Entwicklung von mechatronischen Systemen [9]. Die VDI/VDE 2206:2021-11 „Entwicklung mechatronischer und cyber-physischer Systeme“ hilft bei der Umsetzung solcher Systeme. Sie unterstützt bei der Strukturierung des Entwicklungsprozesses, indem sie detailliert alle Schritte von der Anforderungserhebung bis zur Übergabe an die Nutzung des Systems beschreibt [9].

D. Darstellung der geplanten Vorgehensweise

Aus der vorgestellten Theorie ergeben sich drei übergeordnete Schritte für die Durchführung dieser Masterarbeit, siehe Abbildung 1.

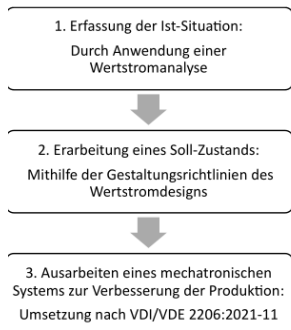


Abbildung 1. Projektierte Vorgehensweise dieser Arbeit.

III. PROZESSIDENTIFIKATION UND KONZEPTIONIERUNG

A. Analyse des Produktionsbereichs

Der zu optimierende Produktionsbereich umfasst die Herstellung von Defibrillationselektroden (Abbildung 2). Das Medizinprodukt ist ein fixer Bestandteil von Defibrillatoren, welche in medizinischen Notfallsituationen als Schockgeber dienen. Die Elektroden werden am Brustkorb der Patient:innen angebracht und werden zur Energieübertragung vom Defibrillator auf den Körper benötigt [10], [11].



Abbildung 2. Defibrillationselektroden inkl. Kabel.

Bei einer ersten Besichtigung des Produktionsbereichs wird festgestellt, dass er in zwei räumlich voneinander getrennte Bereiche aufgeteilt wird:

- „Rohlingsvorbereitung Defi“

- „Montage und Verpackung Defi“

In der Rohlingsvorbereitung werden zehn verschiedene Rohlinge hergestellt. Als Rohling wird die reine Elektrode verstanden, ohne Kabel, Verpackung oder sonstigem. Erst in der Montage und Verpackung werden die Rohlinge mit dem nötigen Zubehör ausgestattet, wodurch sich 18 verschiedene Artikel ergeben. Im Jahr 2022 wurden insgesamt rund 4,5 Millionen Stück von diesen Artikeln verkauft. Bei einem Zweischichtbetrieb mit einer effektiven Nettoarbeitszeit von 7,7h pro Schicht und 249 Produktionstagen, ergibt sich ein Kundentakt von 3,1 s/Stk.

Da für den Bereich „Montage und Verpackung Defi“ seitens des Unternehmens bereits Wertstromanalysen durchgeführt wurden und Verbesserungsprojekte in Planung sind, wird der Fokus auf den Bereich der „Rohlingsvorbereitung Defi“ gelegt.

In diesem Bereich werden zuerst alle Rohlinge auf einer Rohlingsmaschine produziert. Anschließend sind verschiedene Nachbearbeitungstätigkeiten durchzuführen, bevor die Rohlinge an die Montage weitergegeben werden können. Die sehr personalintensiven Nachbearbeitungen bergen ein enormes Verbesserungspotential und werden daher hinsichtlich ihrer Automatisierbarkeit im Rahmen dieser Masterarbeit überprüft.

B. Betrachtung der zu optimierenden Prozesse

Prinzipiell werden bei der Nachbearbeitung zwei Teilprozesse unterschieden.

Beim ersten werden von acht bis zehn Mitarbeiter:innen pro Schicht Restmaterialien entfernt, die aus technologischen Gründen bei der Herstellung am Rohling verbleiben. Je nach Rohlingstyp werden die in Abbildung 3 dargestellten Arbeitsschritte durchgeführt.

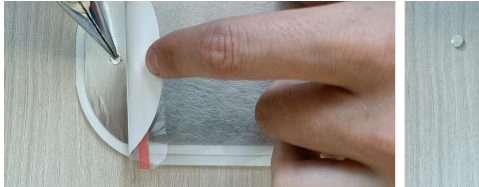
Im zweiten Nachbearbeitungsprozess wird bei einem der Rohlingstypen mithilfe einer manuell betriebenen Stanze ein Nietloch sowie ein Langloch für die Kabelfreistellung erstellt, siehe Abbildung 4.



(a) Entfernen des weißen Schaums/Papiers zum Freilegen der Abziehlasche.



(b) Loch in der Lasche von DFIK freilegen.



(c) Stanzling aus Nietloch entfernen.

Abbildung 3. Bearbeitungsschritte zur Entfernung von überflüssigem Restmaterial.

C. Definition des Umsetzungsprojekts

Die grundsätzliche Anforderung des in dieser Masterarbeit durchgeführten Umsetzungsprojekts ist es, die Machbarkeit einer Maschine zu prüfen, welche die beiden vorgestellten Teilprozesse zu einem gemeinsamen Produktionsprozess vereint. Mithilfe eines Funktionsprototyps sollen die wesentlichen Arbeitsschritte der Nachbearbeitung automatisiert durchgeführt werden. Zudem soll mit dem Testaufbau verifiziert werden, ob der Kundentakt sowie die geforderte Ausschussrate von max. 10% erfüllt werden kann. Die Maschine soll von einer einzelnen Arbeitskraft bedienbar sein und die in Abbildung 5 dargestellten sechs Rohlingstypen bearbeiten können.

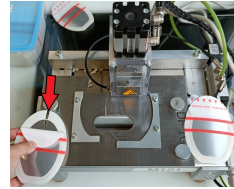


Abbildung 4. Stanzung einer länglichen Kabelfreistellung inkl. Nietloch.



Abbildung 5. Die sechs im Umsetzungsprojekt zu betrachtenden Rohlingstypen.

D. Umsetzungskonzept

Das Umsetzungskonzept basiert auf der Idee, diejenigen Tätigkeiten in den Folgeprozess zu verschieben, die im Vorprozess, also auf der Rohlingsmaschine, aus technologischen Gründen nicht komplett durchgeführt werden können. So wird die Anstanzung des Nietlochs von der Rohlingsmaschine entfernt, da hier der Aufbau der Maschine keine Durchstanzung erlaubt und somit immer eine manuelle Nachbearbeitung notwendig bleibt. Anstelle einer Anstanzung kann die neu zu entwickelnde Maschine das Material, das von der Stanzung verschont werden soll, selbstständig aufklappen. Dadurch wird eine Durchstanzung durch das verbleibende Material ermöglicht und es verbleibt kein zu entfernender

Stanzling im Nietloch.

Somit ändert sich die Bearbeitungstätigkeit vom „Entfernen des Nietloch-Stanzlings“ hin zum „Stanzen des Nietlochs“. Diese Tätigkeit kann durch die Konstruktion einer geeigneten Stanze gleichzeitig mit der Stanzung der Kabelfreistellung erfolgen.

Derselbe Grundgedanke gilt für das Freilegen des Lochs in der Lasche von Rohlingstyp DF1K. Anstatt die Lasche nur anzustanzen, um sie im Nachhinein ausdrücken zu müssen, wird dieser Schritt aus der Rohlingsmaschine entfernt, um mit der neu erarbeiteten Maschine eine Durchstanzung machen zu können.

Die Maschine wird modular aufgebaut. Für jeden Arbeitsschritt wird ein eigenes Modul bzw. eine Station vorgesehen, die von einem Robotersystem angefahren wird. Die Stanzungen werden in einer sogenannten Stanzstation durchgeführt. Eine weitere Station wird für das Entfernen der Restmaterialien vorgesehen. Diese wird als Abziehstation definiert. Mit der Abziehstation sowie mit der Stanzstation werden alle durchzuführenden Tätigkeiten abgedeckt. Zwei Rohlingstypen (DF1C und DF1G) müssen nach der Bearbeitung in der Abziehstation für die Stanzstation gewendet werden. Dazu wird eine sogenannte Wendestation vorgesehen.

Diese drei Stationen decken die kritischsten Arbeitsschritte des zu automatisierenden Produktionsprozesses ab und werden gemeinsam mit einem Roboter auf einem Testtisch schrittweise entwickelt und aufgebaut.

In Abbildung 6 wird das Umsetzungskonzept des Funktionsprototyps gezeigt. Hier ist zu erkennen, dass für die Produktionsmaschine zusätzlich eine Bildverarbeitung sowie ein Start- und Endmagazin geplant sind. Diese Stationen sind für die Überprüfung der Machbarkeit einer auf dem Prototyp basierenden Produktionsmaschine jedoch nicht relevant und werden daher ausgegraut dargestellt.

Um die Bearbeitungsschritte zu testen, reicht im Prototypenstatus ein einzelnes Robotersystem aus. Für die Produktionsmaschine ist jedoch ein zweiter Roboter angedacht, da erwartet wird, dass die Arbeitslast aufzuteilen ist, um den Kundentakt von

3,1 s/Stk erfüllen zu können. Hier wird die Wendestation gleichzeitig als Übergabestation für den zweiten Roboter verwendet.

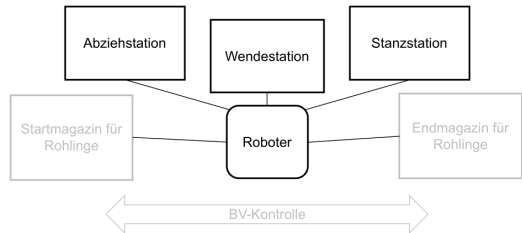


Abbildung 6. Umsetzungskonzept für den Prototyp.

IV. UMSETZUNG UND INBETRIEBNAHME

Nach der Erhebung der Anforderungen sowie dem Entwurf des Systems wird mit dem dritten Schritt nach VDI/VDE 2206:2021-11 fortgefahren, der Implementierung bzw. Umsetzung der Systemelemente. Dabei wird vor allem auf die mechanische Auslegung der einzelnen Stationen und des Roboterwerkzeugs eingegangen.

A. Roboterwerkzeug

Für die Bearbeitungsschritte in den Stationen ist ein Aufklappen der Rohlinge notwendig. Dazu wird das in Abbildung 7 gezeigte Werkzeug entwickelt. Links ist das Roboterwerkzeug im Transportzustand zu sehen. Hier wird mit Vakuumsaugern der Rohling von Station zu Station transportiert. In der rechten Abbildung wird das Roboterwerkzeug mit aufgeklapptem Rohling gezeigt. Für das Aufklappen wird ein Drehzylinder verwendet, auf welchem ein Bauteil montiert ist, das je nach Rohlingstyp mit ein bis drei Saugnapfen ausgestattet ist. Bei einem Rohlingswechsel ist dieses Bauteil zu tauschen. Um die Rüstzeit so kurz wie möglich zu halten, wird das Roboterwerkzeug mit einem Schnellwechselsystem ausgestattet.

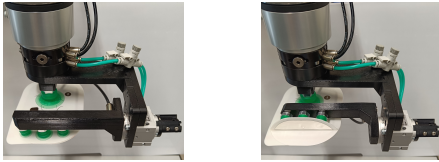


Abbildung 7. Werkzeug des Roboters.

B. Abziehstation

Das Grundprinzip besteht darin, das Restmaterial mithilfe eines Abziehwerkzeugs abzureißen. Der Arbeitsablauf in der Station wird in Abbildung 8 gezeigt. Um das Restmaterial entfernen zu können, muss die oberste Schicht der Rohlinge, also das Trägermaterial, aufgeklappt werden. Dazu befördert der Roboter die Rohlinge zu einem Pneumatikzylinder. Hier wird das zu entfernende Restmaterial niedergedrückt, damit beim Aufklappen des Trägermaterials sichergestellt ist, dass das überflüssige Material nicht ebenfalls aufgeklappt wird. Der aufgeklappte Rohling wird dann zu einem auf einer Linearachse montierten zweiten Zylinder geführt, der mit einem Abziehwerkzeug ausgestattet ist. Der Zylinder drückt das Werkzeug auf den Rohling, während die Linearachse stetig in Abziehrichtung fährt und so das Restmaterial vom Rohling entfernt.

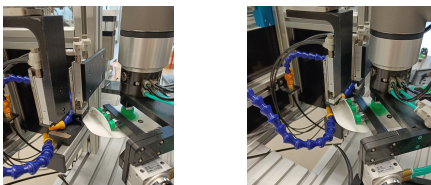


Abbildung 8. Versuchsaufbau der Abziehstation.

C. Wendestation

Die Rohlinge DF1C sowie DF1G müssen für die Bearbeitung in der Stanzstation einmal umgedreht werden. Die dafür notwendige Wendestation kann ebenfalls mit einem pneumatikbetriebenen Drehzylinder umgesetzt werden. Der Aufbau der Station ist in Abbildung 9 abgebildet. Der zu drehende Rohling wird auf einem mit Vakuumsaugern ausgestatteten und mit dem Drehzylinder verbundenen Schwenkarm abgelegt. Dieser dreht den Rohling dann in eine Schablone, damit der Roboter den gewendeten Rohling wieder in der richtigen Position aufnehmen kann. Die Schablonen sind an die entsprechende Rohlingsform angepasst und müssen bei Bedarf schnell gewechselt werden können. Dafür werden die Schablonen in eine Halterung eingeschoben und durch das Anziehen eines Spannzylinders in Position gehalten. Für das Rüsten der Rohlingsschablonen muss somit nur der Spannzylinder pneumatisch entriegelt, die Schablone getauscht und danach der Spannzylinder wieder verriegelt werden.



Abbildung 9. Prototyp der Wendestation.

D. Stanzstation

Die Stanzstation wird in der Entwicklung des Prototyps als unkritisch eingestuft, da das Stanzen von Elektroden in anderen Produktionen des Unternehmens bereits vielfach durchgeführt wird. Trotzdem soll das automatisierte Stanzen im Zusammenspiel mit dem Roboter getestet werden. Dazu wird der in Abbildung 10 gezeigte Aufbau entwickelt. Der Roboter drückt den aufgeklappten Rohling auf eine Auflagefläche und führt ihn in die Stanze ein. Nach

erfolgreicher Stanzen wird er wieder aus der Station entfernt und zugeklappt.

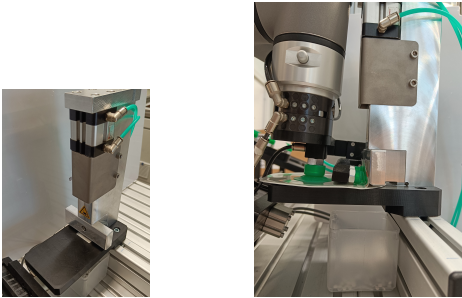


Abbildung 10. Testaufbau der Stanzenstation.

E. Gesamtaufbau

Die einzelnen Stationen werden gemeinsam mit dem Roboter auf einem Testtisch aufgebaut und elektrisch sowie pneumatisch versorgt. Der Maschinenablauf wird auf dem Robotercontroller und einer SPS programmiert. Für eine visuelle Vorstellung des entwickelten Prototyps kann durch das Scannen des in Abbildung 11 dargestellten QR-Codes oder durch das Klicken auf diesen Link ein Video des Maschinenablaufs auf YouTube betrachtet werden.



Abbildung 11. Vorstellung des Prototyps auf Youtube.

V. VALIDIERUNG UND ERGEBNISSE

A. Testablauf

Um den Prototyp zu testen, werden von jedem der sechs Rohlingstypen je zwei Lagerboxen mit 35 Stück bearbeitet. Die Taktzeiten und die auftretenden Fehler werden aufgezeichnet. Außerdem wird

die Rüstzeit beim Wechsel eines Rohlingstyps gemessen. Die Testergebnisse werden mit den vom Unternehmen festgelegten Anforderungen verglichen. Auf Grundlage dessen kann die Machbarkeit einer auf dem Prototyp basierenden Produktionsmaschine diskutiert werden.

B. Prüfergebnisse

Die Taktzeiten t_T für die Bearbeitung aller sechs Rohlingstypen werden in Abbildung 12 visualisiert. Weiters ist der Kundentakt dargestellt, da dieser das zu erreichende Ziel vorgibt.

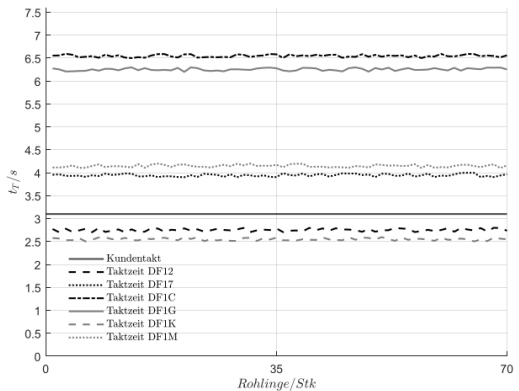


Abbildung 12. Taktzeiten bei der Bearbeitung der Rohlinge.

Die Daten lassen erkennen, dass sich drei Rohlingspaarungen mit ähnlicher Taktzeit ergeben.

Die erste Paarung besteht aus den beiden Rohlingstypen DF1C und DF1G. Beide Taktzeiten befinden sich derzeit klar über dem Soll und über jenen der restlichen Rohlingstypen. Dies liegt daran, dass hier alle Bearbeitungsschritte durchgeführt werden müssen. Für die Einhaltung des Kundentakts von 3,1 s/Stk müssen die beiden Rohlinge mehr als doppelt so schnell bearbeitet werden. Dies bestätigt die Annahme, dass bei der Weiterentwicklung des Prototyps hin zu einer Produktionsmaschine ein zweites

Robotersystem benötigt wird, damit die Arbeitslast aufgeteilt werden kann.

Das nächste auf ähnlicher Taktzeit basierende Paar bildet sich aus den Rohlingen DF17 und DF1M. Auch hier ist die fast idente Taktzeit darauf zurückzuführen, dass die beiden Rohlinge die gleichen Bearbeitungsschritte durchlaufen. Es muss lediglich das Restmaterial von der Abziehlasche entfernt werden. Beide Taktzeiten befinden sich jedoch ebenfalls über dem Kundentakt. Durch eine Optimierung der Abziehstation und der Roboterwege für die Produktionsmaschine sollte dieser aber erfüllbar sein.

Das dritte Paar bilden die Rohlinge DF12 und DF1K. Beide Rohlinge erfüllen bereits die Anforderung des Kundentakts. Ihre Bearbeitungsschritte unterscheiden sich zwar etwas, sind jedoch ähnlich schnell.

Bei der Analyse der auftretenden Fehler wird festgestellt, dass vor allem die Reinigung des Abziehwerkzeugs der Abziehstation optimiert werden muss. Sowohl die Wendestation als auch die Stanzstation arbeiteten bei den Versuchen fehlerfrei. Die Messung der Rüstzeit ergab einen Durchschnittswert von 36,8 s, was die Vorgabe von 60 s deutlich erfüllt.

Zusammenfassend zeigt der Prototypenaufbau, dass eine darauf basierende Produktionsmaschine die Anforderungen erfüllen kann. Es müssen jedoch Optimierungsmaßnahmen ergriffen werden, um die Taktzeiten bei einigen Rohlingstypen und die Ausschussrate der Abziehstation zu verbessern.

VI. ZUSAMMENFASSUNG

Die Produktion für die Herstellung von Defibrillationselektroden stößt an ihre Kapazitätsgrenzen. Durch die Entwicklung neuer Automatisierungslösungen soll hier eine Verbesserung erreicht werden.

Zu diesem Zweck wird der Produktionsbereich mittels einer Wertstromanalyse untersucht. Es werden Verbesserungspotenziale identifiziert und mit der Methodik des Wertstromdesigns ein Zielzustand entwickelt. Dieser Zielzustand soll schrittweise durch

die Umsetzung von Verbesserungsprojekten erreicht werden.

Die Machbarkeit für die Automatisierung der derzeit manuell durchgeführten Arbeitsschritte in der Rohlingsnachbearbeitung soll anhand eines Prototyps getestet werden.

Es wird ein Testtisch mit einem Roboter und mehreren Bearbeitungsstationen aufgebaut. Die erste Station wird für die Entfernung von Restmaterialien benötigt, die aus technologischen Gründen nicht im Vorprozess entfernt werden können. In der zweiten Station wird das automatisierte Stanzen der Rohlinge in Kombination mit einem Roboter getestet. Zwei der sechs zu bearbeitenden Rohlingsformen sind für die Stanzung umzudrehen. Hierfür wird eine Wendestation entwickelt.

Die mit dem Prototyp durchgeführten Tests zeigen, dass die Machbarkeit einer Produktionsmaschine auf Basis des Prototyps gegeben ist. Daher kann mit der Umsetzung einer auf dem Prototyp basierenden Produktionsmaschine begonnen werden.

LITERATUR

- [1] K. Erlach, *Wertstromdesign*. Springer, 2010.
- [2] J. Marchthaler, T. Wigger, and R. Lohe, "Value Management und Wertanalyse," *Wertanalyse—das Tool im Value Management*, pp. 11–38, 2011.
- [3] H. McManus and R. Millard, "Value stream analysis and mapping for product development," 2002.
- [4] B. Singh, S. K. Garg, and S. K. Sharma, "Value stream mapping: literature review and implications for indian industry," *The International Journal of Advanced Manufacturing Technology*, vol. 53, pp. 799–809, 2011.
- [5] K. Koca, "Vergleich von Wertanalyse und Wertstromanalyse als Grundlage zur Optimierung von Prozessen," Ph.D. dissertation, Hochschule für Angewandte Wissenschaften Hamburg, 2017.
- [6] A. Maier, K. Erlach, and M. Ehrat, "Der Kundentakt bestimmt die Montage," 2013.
- [7] M. Veit, "Wertstromorientiertes Prozessmanagement in der Praxis," 2015.
- [8] J. Rießelmann, "Wertstromdesign," 2011.
- [9] *VDI/VDE 2206*. Mess- und Automatisierungstechnik.
- [10] M. Gruner, S. Stegherr, and J. Veith, *Frühdefibrillation*. Stumpf und Kossendey, 2004.
- [11] B. Lüderitz, *Geschichte der Herzrhythmusstörungen: Von der antiken Pulslehre zum implantierbaren Defibrillator*. Springer-Verlag, 2013.

Periodic Boundary Conditions for Non-Uniform Meshes

Dimitri Felbermayr, Mathias H. Luxner (supervisor), and Andreas Mehrle (supervisor)

Abstract—The homogenization method (HM) in combination with the finite element method (FEM) is a powerful technique to obtain thermomechanical material properties out of a microscopic representative volume element (RVE) of composite materials. Therefore, periodic boundary conditions (PBCs) on the finite element (FE) mesh are required, to represent an infinite continuum. Traditionally, PBCs are enforced by coupling the degrees of freedom (DOF) of opposite FE nodes. This limits the conventional method to periodic FE meshes. However, periodic FE meshes are costly and difficult to generate, especially for complicated topologies of RVEs.

In this work, using the programming language C++, two methods are implemented, to enforce PBCs for non-uniform FE meshes of any cuboid RVE. Namely, the radial point interpolation method (RPIM) and the triangulation method (TM). Both methods are used to interpolate the displacement of opposite RVE faces with non-congruent nodal positions.

Using the FEM software ABAQUS, the RPIM and the TM are evaluated and compared. Thereby, it can be shown that both methods are capable of enforcing PBCs on non-uniform FE meshes. Furthermore, the deviation of the homogenized properties compared to the traditional method is negligible. This expands the HM to non-uniform RVE meshes and thus to more efficient and user-friendly mesh generation.

Index Terms—Periodic Boundary Conditions, Non-Uniform Meshes, Homogenization Method, Porous Material, Representative Volume Element, Finite Element Method.

D. Felbermayr is with the Department of Mechatronics, MCI, Innsbruck, Austria, e-mail: d.felbermayr@mci4me.at.

I. INTRODUCTION

HIGHLY tailored materials like composites are becoming increasingly important for lightweight design in mechanical engineering. To optimize the performance of composites for specific requirements, knowledge of their thermomechanical properties is crucial. These properties are dependent on the microgeometry of such materials.

A commonly used method, to investigate the micromechanical material properties, is the homogenization method (HM), which can be used in combination with the finite element method (FEM) [1], [2]. However, this requires periodic boundary conditions (PBCs) on the finite element (FE) mesh.

Traditionally, PBCs are applied to periodic FE meshes by pairwise coupling of opposite FE nodes, see Fig. 1. The requirement of opposite FE nodes severely restricts the application possibilities of HM.

By extending PBCs to non-uniform FE meshes, a broader applicability of the HM is assured. This decouples the micromechanical investigation from the mesh and meshing process. Concerning this goal, previous efforts have already been undertaken by [3]–[7]. In this work, two methods are implemented, to enforce PBCs for an ABAQUS model, having a non-uniform FE mesh. The implementation is done using the programming language C++, whereas PYTHON is used for post-processing.

II. METHODS

A. RVE Homogenization

Based on [1], [2], the effective elastic stiffness properties of orthotropic materials, captured by the tensor of elasticity, can be obtained. This can be done by applying six independent load cases to a 3D representative volume element (RVE), which are strain driven for this work. To gain a link between the macroscopic stress and strain field, $\bar{\sigma}_{ij}$ and $\bar{\epsilon}_{ij}$, and microscopic stress and strain field, σ_{ij} and ϵ_{ij} , the Hill-Mandel lemma [8] is applied. This states that the averaged strain energies at the micro and macro levels have to be equivalent i.e.,

$$\bar{\sigma}_{ij}\bar{\epsilon}_{ij} = \frac{1}{V_{RVE}} \int_{RVE} \sigma_{ij}\epsilon_{ij} dV, \quad (1)$$

where V_{RVE} is the overall volume of the RVE. By discretization, the averaged macroscopic properties can be obtained out of the microscopic properties by

$$\bar{\sigma}_{ij} \approx \frac{1}{V_{RVE}} \sum_{l=1}^N \sigma_{ij}(x, y, z) \Delta V, \quad (2)$$

with the integration point volume ΔV , and

$$\bar{\epsilon}_{ij} \approx \frac{1}{V_{RVE}} \sum_{l=1}^N \epsilon_{ij}(x, y, z) \Delta V. \quad (3)$$

Out of Hook's law [9], the linear relation

$$\bar{\epsilon}_i = \bar{S}_{ij} \bar{\sigma}_j, \quad (4)$$

can be obtained in Voigt notation [2], where \bar{S}_{ij} represents the compliance tensor [9]. The nine independent orthotropic material properties, \bar{E}_i , \bar{G}_{ij} and $\bar{\nu}_{ij}$, can be obtained column-wise out of \bar{S}_{ij} , by applying the six load cases independently [9].

B. Periodic Boundary Conditions (PBCs)

A new strategy to enforce PBCs is developed in this section, based on [10]. For con-

gruent displacements of opposite faces, the coupling is done by eliminating the degrees of freedom (DOF) of each slave node S_i by coupling them to the DOF of corresponding master node M_i with respect to the reference node R_1 . Reference nodes are introduced to apply a predefined displacement to obtain the load cases. In Fig. 1 a symbolic periodic deformation in 2D is made visible. The displacement \vec{u}_{S_i} of the slave node S_i is coupled to the displacement \vec{u}_{M_i} of the master node M_i with respect to the predefined displacement \vec{u}_{R_i} of the reference node R_i , whereas u_{R0} is fully constrained for this work. This leads to a displacement coupling

$$\vec{u}_{S1} = \vec{u}_{M1} + \vec{u}_{R2} - \vec{u}_{R0}, \quad (5)$$

for the left and right edges,

$$\vec{u}_{S2} = \vec{u}_{M2} + \vec{u}_{R1} - \vec{u}_{R0}. \quad (6)$$

for the top and bottom edges and

$$\vec{u}_{S3} = \vec{u}_{R1} + \vec{u}_{R2} - \vec{u}_{R0}. \quad (7)$$

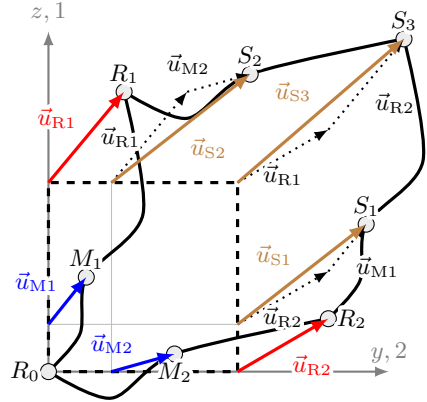


Fig. 1: Symbolic periodic deformation in 2D. Slave nodes (S_i) are coupled to master nodes (M_i) with respect to reference nodes (R_i). Displacement vectors are given by \vec{u} .

for the remaining slave vertex. To expand this method to the 3D case, designations for vertices, edges and faces are introduced in Fig. 2. Every slave node can only be coupled once [11]. For a cuboid, a different number of vertices, edges, and faces is existing. Therefore, the coupling of vertex-, edge- and face- nodes has to be done differently, to avoid a over-constrained model. The coupling of vertices is done according to Tab. I. Each nodal coupling of all DOF (three for solid elements) is designated by \times .

The DOF coupling of the edge nodes can be done according to Tab. II, whereas the DOF coupling of face nodes is done according to Tab. III. Taking all those coupling constraints into account, PBCs are given for a RVE having a periodic FE mesh.

However, the application of PBCs to non-uniform FE meshes, for which the surfaces of periodicity can have an arbitrary mesh, demands more advanced treatment. This can be fulfilled by interpolation methods, discussed in the following sections.

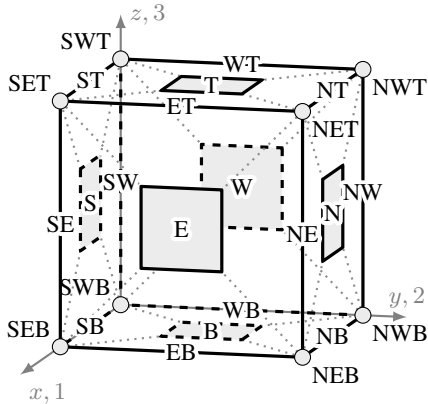


Fig. 2: Celestial RVE notation of vertices, edges and faces.

Tab. I: DOF-coupling of vertex nodes with respect to reference nodes. \times represents a coupling of all nodal DOF.

Slave	Reference			
	R ₀	R _{x,1}	R _{y,2}	R _{z,3}
SWB	\times	-	-	-
SEB	\times	\times	-	-
NWB	\times	-	\times	-
SWT	\times	-	-	\times
NEB	\times	\times	\times	-
SET	\times	\times	-	\times
NET	\times	\times	\times	\times
NWT	\times	-	\times	\times

Tab. II: DOF-coupling of edge nodes with respect to reference nodes. \times represents a coupling of all nodal DOF.

Slave	Reference				Master		
	R ₀	R _{x,1}	R _{y,2}	R _{z,3}	SB	WB	SW
EB	\times	\times	-	-	-	\times	-
NB	\times	-	\times	-	\times	-	-
SE	\times	\times	-	-	-	-	\times
NE	\times	\times	\times	-	-	-	\times
NW	\times	-	\times	-	-	-	\times
ST	\times	-	-	\times	\times	-	-
ET	\times	\times	-	\times	-	\times	-
NT	\times	-	\times	\times	\times	-	-
WT	\times	-	-	\times	-	\times	-

Tab. III: DOF-coupling of face nodes with respect to reference nodes. \times represents a coupling of all nodal DOF.

Slave	Reference				Master		
	R ₀	R _{x,1}	R _{y,2}	R _{z,3}	B	S	W
E	\times	\times	-	-	-	-	\times
N	\times	-	\times	-	-	\times	-
T	\times	-	-	\times	\times	-	-

C. Radial Point Interpolation Method (RPIM)

The radial point interpolation method (RPIM) is a numerical interpolation technique. Based on [12], where it is used for

meshfree methods, the theoretical description is carried out. The basic idea is, that given nodes in the domain of the master face Ω_M are selected to interpolate the displacement of a node S in the domain of the slave face Ω_S . A schematic of the RPIM is given in Fig. 3. The size and shape of the domain partition in Ω_M and thus the number n of included nodes for the interpolation can be chosen freely. To interpolate a displacement

$$u_S = \sum_{i=1}^n R_i a_i + \sum_{j=1}^m p_j b_j \quad (8)$$

of a node S , the RPIM can use a set of n scattered nodes in Ω_M , where R_i and p_j represent the radial basis function (RBF) distance weight, and the polynomial term with the order m , in Ω_S respectively. R and p_j can be obtained with regard to [12]. The coefficients which have to be obtained from the master nodes are given by a_i and b_j . A reformulation, according to [12], leads to the displacement

$$u_S = \begin{pmatrix} \vec{R}_S^T & \vec{P}_S^T \end{pmatrix} \mathbf{G}_M^{-1} \vec{U}_M, \quad (9)$$

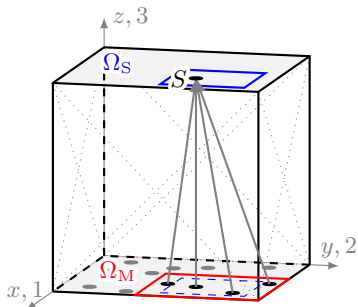


Fig. 3: RPIM interpolation of node S in the domain partition of the slave face Ω_S by involving several nodes in the domain partition of the master face Ω_M .

in matrix notation, where

$$\mathbf{G}_M = \begin{pmatrix} \mathbf{R}_M & \mathbf{P}_M \\ \mathbf{P}_M^T & \mathbf{0} \end{pmatrix}, \quad (10)$$

contains the RBF matrix \mathbf{R}_M and the polynomial moment matrix (PMM) \mathbf{P}_M in Ω_M , whereas \vec{U}_M represents the displacement magnitude of n master nodes in Ω_M , which is not known beforehand. More detailed information is given in [12]. However, this leads to the coefficient vector

$$\vec{\Phi}^T = \begin{pmatrix} \vec{R}_S^T & \vec{P}_S^T \end{pmatrix} \mathbf{G}_M^{-1}, \quad (11)$$

containing the coefficients which are assigned to the nodes in Ω_M . For this implementation, a local partitioning of Ω_S and Ω_M is done by partitioning each face in several rectangular domains like it can be seen in Fig. 3. In order to avoid overshooting behavior, the interpolation at the boundaries of the scattered nodes is avoided. Therefore, the domain partition in Ω_M is chosen to be larger by the length $d_{\text{dom-part}}$ in every direction than the partition in Ω_S . This can also be seen in Fig. 3. For this implementation, five parameters can be set for individual applications of the RPIM. These are provided in Tab. IV. The parameters α_c and q have to be set, to gain the RBFs, whereas m can be set to gain the PMMs [12].

D. Triangulation Method (TM)

The derivation of the triangulation method (TM) is based on [13]. The basic idea is,

Tab. IV: Parameters, which can be selected for the RPIM interpolation.

Parameter	Description	Unit
$n_{\text{dom-part}}$	Partitions in Ω_M and Ω_S	-
$d_{\text{dom-part}}$	Enlargement of Ω_M	mm
α_c	Shape parameter	-
q	Shape parameter	-
m	Polynomial order	-

that the displacement u_S of a node S in the domain of the slave face Ω_S is interpolated by using the displacement magnitude \vec{U}_M of three nodes, M_A , M_B and M_C , building a triangle in the domain of the master face Ω_M . A schematic representation can be seen in Fig. 4. Each node of Ω_S is uniquely part of one triangle projection of Ω_M if a tetrahedral mesh is considered. The displacement

$$u_S = \sum_{i=1}^3 \phi_i U_{Mi} \quad (12)$$

therefore can be obtained, by gaining the weights ϕ_i . This can be done, by partitioning each face of a triangle, with regard to the node S . In Fig. 5 this is visualized. A_A , A_B and A_C represent the area partitions, obtained by the edge lengths a , b and c , which sum up to the overall area A of the triangle. Therefore, each single weight

$$\phi_i = \frac{A_i}{A}, \quad (13)$$

can be obtained which leads to the interpolation of the displacement

$$u_S = \vec{\Phi}^T \vec{U}_M. \quad (14)$$

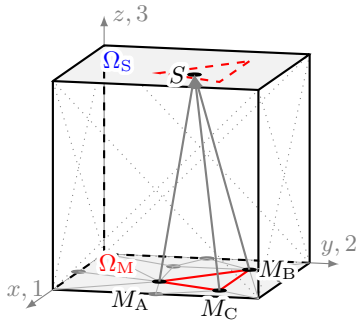


Fig. 4: TM interpolation of node S on the slave face Ω_S by using three nodes, M_A , M_B and M_C , of the master domain Ω_M .

of one slave node. By using ABAQUS C3D10 elements, each element comes up having ten nodes. These are four edge nodes and six mid-nodes. The node numbering can be seen in Fig. 6. It can be further seen, that each face of the C3D10 elements can be partitioned into four triangles, by an orientation on the node numbering. This is used in this work to increase the interpolation accuracy of the TM.

E. C++ Implementation

The implementation of the RPIM and the TM is done using the programming language C++. An overall flow chart can be seen in Fig. 7. The modular structure enables, to choose either the RPIM or the TM for interpolation. The main steps are

- (i) **Import mesh:** In the first step, the mesh is imported from a given ABAQUS in-

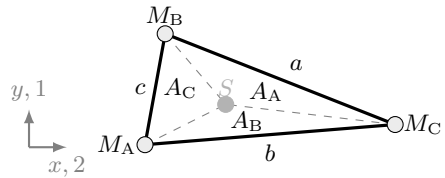


Fig. 5: Partitioning of a triangle, defined by M_A , M_B and M_C , into three area fractions according to S .

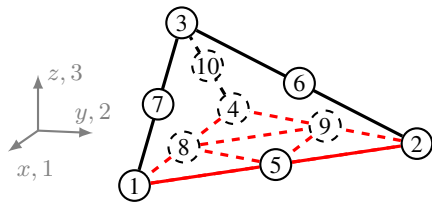


Fig. 6: Partitioning of an ABAQUS C3D10 element face into four triangles [11].

put file, containing the nodes and elements of a RVE model.

- (ii) **Sort nodes and elements:** The nodes and elements, which are part of the RVE boundary, are sorted in a way, that they are clearly arranged and accessible. This is done according to the designations given in Fig. 2.
- (iii) **Interpolate:** Then the method can be selected. Therefore, the interpolation is either done using the RPIM or the TM.
- (iv) **Apply PBCs:** In the last step, PBCs are applied by generating constraint equations which are applied by opening the model with ABAQUS/CAE [11].

PYTHON is used for post-processing.

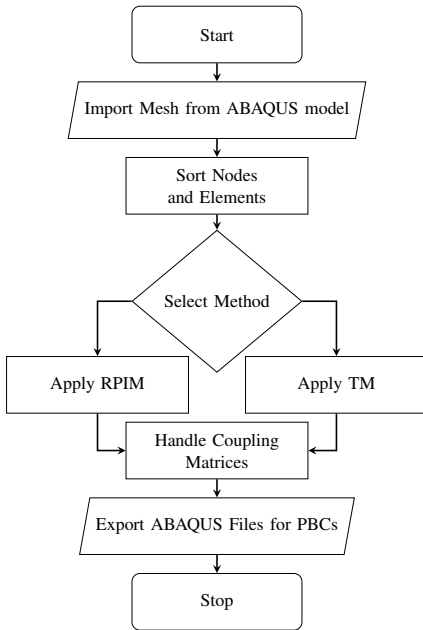


Fig. 7: Flow chart of the overall process to apply PBCs to a given ABAQUS model using the RPIM or the TM for interpolation..

III. RESULTS AND INTERPRETATION

Two different models are used for the evaluation. This is on one hand a homogeneous isotropic RVE and on the other hand a porous RVE, which is a commonly practical application of the HM.

A. Homogeneous Isotropic RVE

For the homogeneous isotropic RVE a mesh convergence study is carried out using four different refinement factors, namely $C_r = 1$, $C_r = 2$, $C_r = 4$ and $C_r = 8$. The RPIM parameters, which are chosen therefore are given in Tab. V. The FEM response of the model with the mesh refinement factor $C_r = 1$ is visualized in Fig. 8 for the RPIM. A similar response is obtained by using the TM and therefore not visualized. For both methods, the RPIM and the TM, a convergent behavior regarding the homogenized properties is given. This can be seen in Fig. 9. Even for the mesh refinement factor $C_r = 1$, which is a coarse mesh, the average deviation of the effective elastic material properties e_{mp} is below 1.5% for both methods.

B. Porous RVE

The simulation of the porous RVE is done using the parameters specified in Tab. VI.

Tab. V: Selected RPIM parameters for interpolating displacements with regard to the mesh refinement factor C_r .

Parameter	Magnitude				Unit
C_r	1	2	4	8	-
$n_{dom-part}$	2	3	10	30	-
$d_{dom-part}$	2	1.8	1	0.4	mm
α_c	2	6	8	10	-
q	0.98	0.98	0.98	0.98	-
m	4	6	6	8	-

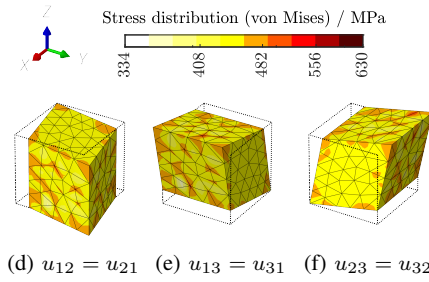
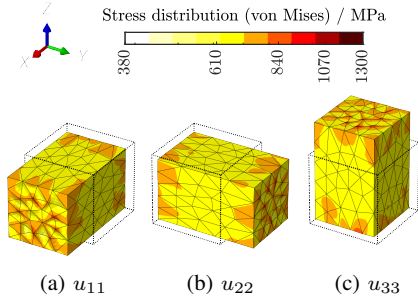


Fig. 8: Homogeneous isotropic RVE response, (a), (b) and (c) represent the tensile load cases and (d), (e) and (f) the pure shear load cases with a displacement factor of 40.

Tab. VI: Selected RPIM parameters for interpolating displacements of the porous RVE.

Parameter	Magnitude	Unit
$n_{\text{dom-part}}$	30	-
$d_{\text{dom-part}}$	0.5	mm
α_c	10	-
q	0.98	-
m	4	-

The response of the porous RVE can be seen in Fig. 10 with respect to the six independent load cases by using the TM. A similar response can be observed when using RPIM for interpolation. The evaluation is done, by comparing the homogenized properties, obtained by applying the RPIM and the TM,

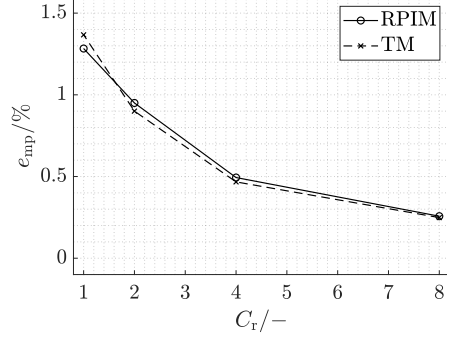


Fig. 9: Effective elastic material property percentage error e_{mp} with respect to the mesh refinement factor C_r .

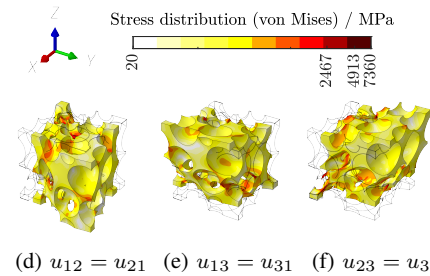
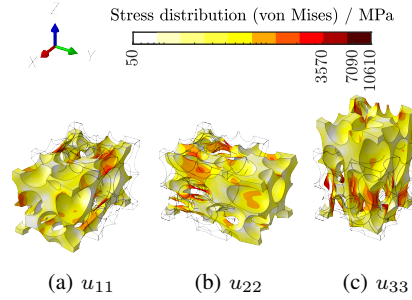


Fig. 10: TM response of the porous RVE, whereas (a), (b) and (c) represent the tensile load cases and (d), (e) and (f) the pure shear load cases with a displacement factor of 40.

to the periodic solution. This reveals an average deviation of the homogenized material properties of $e_{mp} = 0.0161\%$ for the RPIM and $e_{mp} = 0.0124\%$ for the TM.

IV. CONCLUSION

Both methods, the RPIM as well as the TM, can be used in combination with the HM to obtain the homogenized properties of any cuboid RVE having a non-uniform mesh. The deviation, even for coarse meshes, is small. This can be seen in Fig. 9. The computational effort is higher for the RPIM compared to the TM. The reason for that is, that the RPIM, at least for this implementation, includes more nodes in Ω_M for the interpolation of a slave node S . It has to be mentioned, that at the boundaries of the RVE, stress concentrations are occurring because of the coupling. However, these have no big impact on the homogenization procedure.

ACKNOWLEDGMENT

Special thanks to the company LUXNER Engineering ZT GmbH for initiating this project, and the MCI for being an inspiring place to conduct a master's program.

REFERENCES

- [1] J. Yvonnet, *Computational homogenization of heterogeneous materials with finite elements*. Springer, 2019, vol. 258.
- [2] H. J. Böhm, "A short introduction to basic aspects of continuum micromechanics," *Cdf-fmd Report*, vol. 3, 1998.
- [3] V.-D. Nguyen, E. Béchet, C. Geuzaine, and L. Noels, "Imposing periodic boundary condition on arbitrary meshes by polynomial interpolation," *Computational Materials Science*, vol. 55, pp. 390–406, 2012.
- [4] A. Akpoyomare, M. Okereke, and M. Bingley, "Virtual testing of composites: Imposing periodic boundary conditions on general finite element meshes," *Composite Structures*, vol. 160, pp. 983–994, 2017.

- [5] F. Reis and F. A. Pires, "A mortar based approach for the enforcement of periodic boundary conditions on arbitrarily generated meshes," *Computer Methods in Applied Mechanics and Engineering*, vol. 274, pp. 168–191, 2014.
- [6] J. Tyrus, M. Gosz, and E. DeSantiago, "A local finite element implementation for imposing periodic boundary conditions on composite micromechanical models," *International Journal of Solids and Structures*, vol. 44, no. 9, pp. 2972–2989, 2007.
- [7] R. Wang, L. Zhang, D. Hu, C. Liu, X. Shen, C. Cho, and B. Li, "A novel approach to impose periodic boundary condition on braided composite rve model based on rpim," *Composite Structures*, vol. 163, pp. 77–88, 2017.
- [8] F. Nicot, N. P. Kruyt, and O. Millet, "On hill's lemma in continuum mechanics," *Acta mechanica*, vol. 228, pp. 1581–1596, 2017.
- [9] H. A. Mang and G. Hofstetter, *Festigkeitslehre*. Springer-Verlag, 2013.
- [10] T. Daxner, "Multi-scale modeling and simulation of metallic foams," Ph.D. dissertation, Technische Universität Wien, 2002.
- [11] M. Smith, *ABAQUS/Standard User's Manual, Version 6.9*. United States: Dassault Systèmes Simulia Corp, 2009.
- [12] G.-R. Liu and Y.-T. Gu, *An introduction to meshfree methods and their programming*. Springer Science & Business Media, 2005.
- [13] M. S. Floater, "Generalized barycentric coordinates and applications," *Acta Numerica*, vol. 24, pp. 161–214, 2015.



Dimitri Felbermayr is a Master's student at the mechatronics department at the MCI, majoring in mechanical engineering. In his bachelor thesis, he dealt with acoustic properties of cellular materials at the FH-Kärnten.

Data Structure for Virtual Commissioning - Exchange Relevant Information Focusing on Integration of Modular Component Models and Wide Compatibility

Dominique Geiger, B.Sc., Benjamin Massow, B.Sc., M.Sc. (supervisor) and
Thomas Hausberger, B.Sc., M.Sc. (supervisor)

Abstract—This paper presents a data structure consisting of CAD assemblies, physical behaviour models, example control code and documentation used for a virtual commissioning.

Virtual commissioning is becoming increasingly important in the engineering process of plants and production machines. This increase results from time and thus also cost savings during the entire product development and especially during commissioning as outcome of a virtual commissioning. However, this process also has a drawback in form of an increased effort in the modelling of a virtual plant and the complexity associated with it. The behaviour modelling of different components can be done in several different tools, mostly depending on the manufacturer. The CAD data are also available in different file formats and are thus no exception of this rule. This variety of possibilities increases the complexity of the integration process enormously. However, for virtual commissioning to be efficient, the integration of the used components should be as simple as it could be.

As a possible solution to this problem, this paper proposes an exemplary data structure to simplify data exchange. This structure consists of the necessary information for a virtual commissioning such as CAD assemblies with kinematization, physical behaviour model, control code and related documentation. The information exchange is done by means of standardised and common interfaces to be able to cover a wide range of possible application areas. Finally, the usage and the resulting benefits of this data structure will be demonstrated with the help of a practical example. In the example shown, the behaviour model and control are

executed on two different machines ensuring real-time capability and forming a hardware in the loop (HIL) system. The behaviour models of different subsystems are generated in Simulink and integrated into the PLC run-time using TcCom objects in TwinCAT. The creation of control code for the plant is greatly simplified using ready-to-use code snippets, provided by suppliers in this case.

With the help of the proposed data structure, the process of virtual commissioning can be simplified and thus accelerated by a reduction in complexity. No further expertise is required to use the data structure besides the basic knowledge of PLC programming, handling of CAD software and the modelling of systems for example in Simulink. However, there is still room for improvement especially in the exchange of CAD assemblies including a kinematization and the integration of behaviour models in a PLC.

Index Terms—Virtual Commissioning, PLC, Automation, Modularity, Data Structure.

I. INTRODUCTION

VIRTUAL commissioning is gaining relevance in many industry branches due to its advantage of reducing time and therefore costs. However, the creation and especially the integration of plant models is a hurdle. A unified data structure consisting of the geometry, behaviour model and exemplary PLC code can reduce this issue.

Aim of this paper is to develop a data structure consisting of kinematized CAD assemblies, physical behaviour models and PLC code for an easier

D. Geiger studies at MCI Innsbruck, e-mail: dm.geiger@mci4me.at.

integration in a virtual commissioning. With this data structure, the exchange of components between supplier and customer can be standardized and, in addition, the setup of a virtual commissioning can be simplified.

II. STATE OF THE ART

A. Examples for a Virtual Commissioning

A virtual commissioning can be done in different ways, whereby the chosen method also strongly depends on the specific application. In this chapter, various possibilities will be briefly explained.

1) *Visualization in the PLC Development Environment*: The human-machine interface (HMI) offers a quick possibility for testing the control software. In most cases, a graphical user interface (GUI) can be created directly in the development environment in which current values and outputs of the control software can be displayed and inputs can be set. In this approach, the human developer imitates the behaviour of the real plant and therefore tries to identify potential problems in the software. Due to the human interaction, fast processes and reactions can only be tested to a limited extent. However, basic code sections and functions can be evaluated during development and in the process of a virtual commissioning.

2) *Plant Model using a Physics Engine*: The next step is the use of an external physics engine such as Unity. Especially by using existing libraries and well documented interfaces a simple model can be created in a short time. In addition, the handling of the virtual plant can be simplified by using the original geometry data. A disadvantage of this method compared to the direct visualization in the PLC development environment is the requirement of an additional communication between the PLC run-time and the physics engine. Nevertheless, depending on the implementation, this procedure can also be carried out in real-time. By using the original CAD data, a visually appealing model can be created and controlled via the PLC. This is an important advantage especially in presentations with customers

or in sales, in order to be able to address people from the non-technical area as well. An example for the implementation in Unity and TwinCAT is [1]. Here, a digital twin is tested and used in a HIL system, achieving a communication time of 10 ms and a step time of the simulation in Unity of 5 μ s.

3) *Digital Twin in a Simulation Software*: The use of simulation software such as MATLAB/Simulink is particularly advantageous for complex systems or in control engineering, as these systems can be easily created using mathematical equations. Furthermore, the equations and a GUI improve the overview and reduce time needed for maintenance. In this method, the run-time of the virtual plant is executed in the simulation software and for this reason also requires additional communication to the PLC run-time. If the system is supposed to be real-time capable, special hardware or precautions are often required.

B. Required Information for Setup

For the identification of the required information of a virtual commissioning, the general process of product development has to be examined. The development of a product includes several stages in different engineering disciplines. These stages can partly be developed in parallel to save time and costs in the development process, but special attention must be paid to the dependencies between them. In general, the development process can be summarized into following steps: Process design, mechanical and electrical engineering, software development and commissioning. The first phase of product development includes the general boundary conditions like defining objectives, constraints and interfaces. The next step is to develop the product's hardware usually consisting of mechanical and electrical components. The hardware must be developed in collaboration between both disciplines, since changes in the mechanics, for example, often also lead to changes in the electrics. Parallel to the development of the hardware, the process of software development can be started already. For the completion of the software, however, the hardware must already have been

finalized. The final step in product development is an optional virtual commissioning followed by actual commissioning.

Information from all phases of product development are required for the successful execution of a virtual or a real commissioning:

1) *Process Design*: Interfaces to other parts of the plant, such as transfer points of raw and finished parts, but also maximum time limits, throughput quantities and the selection and characteristic of used sensors and actuators belong to this area. The characteristic of the used hardware is particularly important in the creation of behaviour models, where the information is given in data sheets or already integrated in models.

2) *Mechanical Engineering*: The general structure of the mechanics and kinematization of the individual components are defined in this step. The design of the product and the used materials determine the physical behaviour and results in multiple CAD files.

3) *Electrical Engineering*: The configuration of the PLC with the used terminals is part of this section. Especially interesting for the development of the software is the mapping between the inputs and outputs of the terminals with the connected devices. The documentation of the mapping and the structure of the PLC can be shown in a tabular form or in a dedicated object in the development environment of the PLC.

4) *PLC Coding*: If more complex components are used in the plant, often these are addressed via their own interface. The documentation and the knowledge of handling these interfaces is important for the creation of the software but also for the later commissioning. In the best case examples exist, which explain the handling and therefore ensure a faster implementation. But also, an example PLC code of simpler components can often be advantageous.

C. Criteria for a Data Structure

Based on these fundamentals of a virtual commissioning, criteria for a data structure can be identified. With the help of these criteria, the efficiency of a data

structure can be determined, whereby these should be fulfilled as good as possible. The identified criteria are:

- Independent of used hardware setup (HIL/SIL).
- Represent relevant information.
- Modular layout in order to represent components of a real plants.
- High compatibility with common tools.
- Use of internationally standard data formats.
- Low level of additional knowledge required.

III. PROPOSED DATA STRUCTURE

Based on the previous findings, a data structure is developed and presented in this chapter. The main objective of this data structure is a high compatibility and a modular design in order to be able to represent the individual sub-components of a real plant. The data structure itself should contain all necessary information required for a virtual commissioning.

A. Layout of the Data Structure

The actual layout of the data structure is kept simple and consists of a root folder with several directories for the relevant information as shown in Figure 1. The sub-directories contain the information of kinematization and geometry, physical behaviour model, PLC source code and documentation. An optional ReadMe.txt file used for general information and revision control of the structure itself. Finally, the root folder is compressed into a .zip file for an easier distribution. Due to the plain and simple structure, it can be easily extended and modular exchanged.

B. Selected File Formats

With the layout of the data structure appropriate data formats for representation of CAD assemblies, physical behaviour models and PLC code are now selected. For this purpose different possibilities of implementation are tested on the basis of the used software shown in Table I.

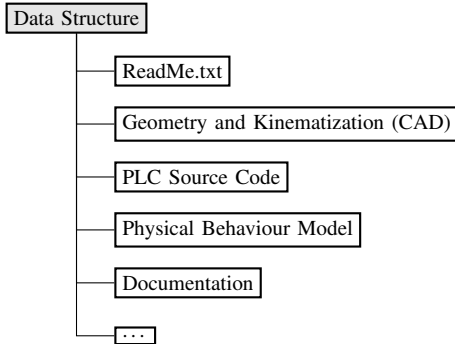


Fig. 1: Layout of proposed data structure containing required information for a virtual commissioning.

TABLE I: Used software for selection of suitable data formats of the data structure.

Area	Used Software
CAD	Autodesk Inventor Professional 2022 Autodesk 3ds Max 2022
PLC	TwinCAT 3 Version 3.1.4024.25
Modelling	MATLAB R2020b, Simulink

1) *CAD Data*: The exchange of CAD data between different tools is usually done via neutral formats such as *.step* or *.iges*. However, only the pure geometry is supported and transferred via these formats and an information loss of the kinematization occurs. In many areas and also in virtual commissioning, kinematization and thus its exchange is, however, an important prerequisite. The industry has recognized this problem of the missing interface of the kinematization and is therefore working on different solutions. For example, existing CAD data formats can be extended to support features and kinematization. For this purpose, [2] describes a possible extension of the *.step* format in order to include the kinematization. A second possibility is the development of a new and neutral data format. In order to be able to use this format, an integration

into existing CAD tools can be done or, alternatively, a conversion of the native formats with the help of translators is also an option. However, both methods are feasible but require a major effort in technical and organizational terms. The conversion of native features in the design of a part into a neutral format and back into a second CAD software is shown for example in [3]. The same approach could be used to export kinematization as well as CAD features, as shown in [4]. As an alternative, the kinematization could be saved in an additional file and linked to the original CAD data as shown for example in [5]. A promising solution is the COLLADA format, which supports kinematization starting from version 1.5 [6]. The COLLADA format, released back in 2004, is based on *.xml* documents and is mainly used in the entertainment and gaming industry suffering from the same problem with a large number of incompatible tools. The use in the manufacturing industry is currently not attractive, because suitable tools for the conversion to and from native formats are still missing. Since there is no established solution for exporting kinematic CAD assemblies, there are two options for the data structure: Either a neutral format such as *.step* is used, which has a high compatibility but does not support kinematization, or a native format is used, in which compatibility is reduced but kinematization is preserved.

2) *Behaviour Model*: Similar to the geometry data, there is a variety of possible file formats for the description of a physical behaviour model. This is mostly dependent on the discipline and the simulation software used. However, especially in the field of co-simulation a universal interface is required to simplify data exchange due to the combination of multiple engineering disciplines. For this reason, the Functional Mockup Interface (FMI) was defined by Modelica Association already in 2010 and is currently in fact the default format for exchanging models. Currently this interface is available in version 3 and is already supported by more than 170 different tools [7]. For this reason, the FMI format is selected to represent physical models in *.fmu* files.

3) *PLC Code*: In the area of control engineering and automation using PLCs, the basis is mainly the international IEC 61131 standard. This standard is based to a large extent on the organization PLCopen, which has set itself the goal of increasing the efficiency in the creation of control software and to be platform-independent between different development environments. To make this possible the PLC project with its code and libraries is saved in .xml files and therefore forming an universal interface. This exchange format was standardized in 2019 in part 10 of the IEC 61131 standard. Thanks to the international validity, many manufacturers rely on this standardization and offer converters to and from the PLCopen format.

4) *Overview of Selected Data Formats*: Based on these findings an overview of the selected data formats is found in Table II.

TABLE II: Chosen file formats for the data structure.

Information	File Format
CAD as pure geometry	Neutral like .step
CAD with kinematization	Native like .iam or COLLADA
Physical model	FMI as .fmu
PLC code	PLCopen as .xml

C. General Workflow of Using the Data Structure

The workflow when using the proposed data structure is shown in Figure 2 and consists of two main parts: Generating the structure (export) and using the structure for example in a virtual commissioning (import). The export of data into the structure is more straightforward and thus easier to accomplish compared to the import. The data structure can usually be generated in just a few steps, with CAD data being the exception to this general rule. In principle, the greatest effort is required when using the data structure in the field of CAD data, due to the lack of an interface for saving the kinematization of an assembly. If only the geometry should or can be

transferred, a neutral format like .step or .iges is recommended, which are supported by most of the CAD tools. Until the COLLADA format or alternatives are established in industry, exporting kinematization is done easiest in native CAD formats. In comparison, the export of the behaviour model, the control code of the PLC and, if necessary, a technical documentation do not cause serious problems and are supported by the majority of available software.

The process of importing the data structure is more complex, where the individual steps strongly depend on the used target system. In the case of CAD data, it is necessary to distinguish between three possible approaches: the assembly is not kinematized (for example .step), the assembly is kinematized in neutral format (for example COLLADA) and the assembly is kinematized in a native format (for example .iam). The native format provides the least effort for integration, followed by plain geometry without kinematization (depending on the complexity of the assembly). The greatest effort represents the import of a COLLADA file, due to the lack of conversion tools. The import of the PLC code is done in the majority of PLC systems without any difficulties thanks to the international standardization of the programming language. As a result sample code can be integrated into an existing project fast and without any difficulties. However, with the behaviour model, a dependency to the used software is once again recognizable. In the best case, the .fmu object can already be executed directly in the PLC run-time and linked to the inputs and outputs of the hardware. Alternatively, the model can be executed in a simulation software, which again communicates with the PLC run-time.

IV. EXEMPLARY APPLICATION

The practical use and individual steps in the implementation of the proposed data structure are shown by means of an example using the same software as shown in Table I. As example of a real plant a modular Teaching Factory for automation technology and PLC is used as shown in Figure 3.

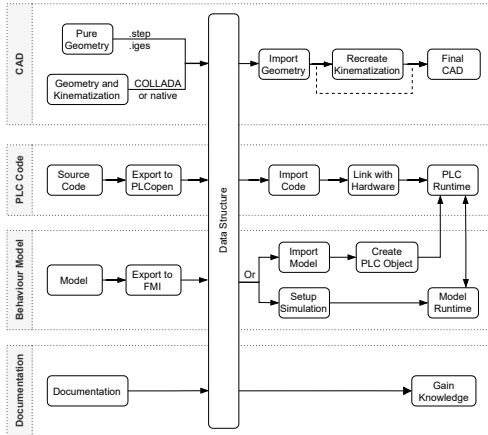


Fig. 2: General workflow of proposed data structure consisting of the most important steps for export and import.

A. Generate the Data Structure

The first step is the generation of the proposed data structure consisting of CAD assemblies, behaviour model, PLC source code and optional documentation. This data export is done from the view of a supplier for each module of the plant.

For exporting the CAD data, this example uses the native format of Inventor for kinematized assemblies. This assumes that supplier (transmitter) and customer (receiver) of the data structure both use Inventor. If the assembly is purely static and thus no kinematization is required, the .step format is used for exchange. In addition to the .fmu files, the behaviour models are included also as original Simulink files into the data structure. This is done because the required product TE1420 to integrate the .fmu models into TwinCAT is still under development [8]. For this reason, the models are integrated into TwinCAT via code generation from Simulink as alternative to the direct way using TE1420 [9]. The intended product, as well as the alternative generate a TcCom object, which is then linked to the inputs and outputs of

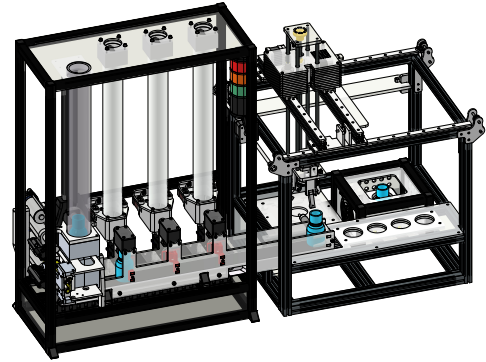


Fig. 3: The used Teaching Factory representing a filling station. On the left is the module Separation followed by the Conveyor Belt with three Dosing Units. The right half contains the Cartesian Gripper with a Load Cell (left), Thermal Processing (rear right) and an output storage (front).

the PLC. Therefore, the result of both methods is identical, but the source format is different. The interface of the models reproduces the signals of the real hardware and describes the given system. The level of detail of the models can vary depending on the requirements of the simulation. In the end the source code of the PLC is exported to the PLCopen format for exchange.

The collected information is finally bundled in the data structure as shown for an exemplary module in Figure 4.

B. Import the Data Structure

The integration of the data structure in a virtual commissioning is also done for the areas: CAD assembly, behaviour model and PLC code. In the case of CAD data, the integration is a standard process for an average user of Inventor. If the kinematization is lost during the export, for example when only the pure geometry is exchanged, it has to be recreated by the customer. The integration of PLC code into TwinCAT is similarly simple due to the standardization

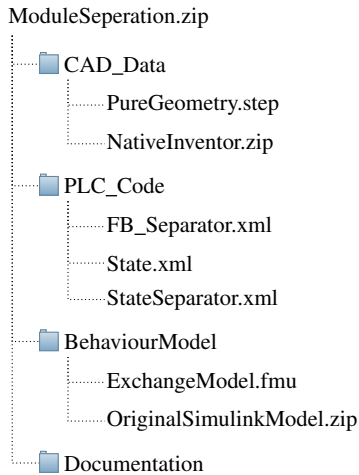


Fig. 4: Data structure of module Separation consisting of CAD data as pure geometry and including kinematization in a native format, PLC code and the behaviour model as .fmu and original Simulink files.

of the PLCopen format. As already mentioned, the product for integrating the behaviour models is under development and therefore code generation is used as an alternative. For this purpose, the original model from Simulink is used and converted into a TcCom object in TwinCAT [10]. In this way, a separate object is created on the PLC for each module, which is finally connected to the inputs and outputs of the control code. With the integrated behaviour models and PLC code, a virtual commissioning can now be performed. In this example the model of the plant and the control code are executed on two different machines therefore forming a HIL system with a real-time communication between them. Using the data structure in a software in the loop (SIL) system is also possible.

V. RESULTS AND EVALUATION

The result of this paper is the proposed data structure bundling information for a virtual commis-

sioning. The structure is modular designed and data is stored in compatible formats. The practical use of this data structure is evaluated by means of an example shown in section IV using the criteria from subsection II-C.

Through the evaluation it is stated that the data structure can basically be used independent of the hardware setup for example in a HIL system. The data required for virtual commissioning, such as CAD assemblies, behaviour models, PLC code and documentation, can be represented in the data structure without problems. As shown in the example, individual modules of a plant can be reproduced with the data structure and finally integrated into a higher-level system. However, there is still room for development in the compatibility of the data formats, especially in the area of kinematized CAD data and in the integration of .fmu models in a PLC project. While the integration of behaviour models in the case of Beckhoff is only a matter of time, the exchange of kinematization in CAD data represents a bigger challenge. In this case, a practicable solution has not yet been established in the industry. In terms of the requirements for the additional know-how needed, no further skills are required in the case of the CAD data and the pure PLC code. When creating the behaviour models, basic knowledge of programming a PLC is an advantage but not absolutely necessary. However, the integration of the behaviour models into the PLC project requires additional knowledge, which can be acquired in a seminar, for example. The result of this evaluation is summarized in Table III and is valid for the used software versions shown in Table I, where upward compatibility is likely but downward compatibility is not guaranteed.

VI. SUMMARY

In the scope of this paper, a data structure for the exchange of relevant information for a virtual commissioning is developed. This structure is focused on high compatibility between different tools and has a modular structure. The covered information of the data structure includes the mechanical design

TABLE III: Evaluation of the data structure with respect to the shown example from bad (○○○) to good (●●●). Although the defined information is represented in the data structure, compatibility could be further improved.

Criterion	Evaluation
Independent of hardware setup (HIL/SIL)	●●●
Represent relevant information	●●●
Modular layout to represent components	●●●
High compatibility	●○○
Using standard formats	●●●
Low level of additional knowledge	●●○

represented by the CAD data, physical behaviour models of different used components and source code for the control via a PLC. Additional information such as data sheets can be added to the data structure as required. The presented data structure is evaluated on the basis of an example for its usability. The result of this evaluation shows potential for improvement, especially in the exchange of kinematized CAD assemblies. Besides the native formats there is no established solution available at the moment. Furthermore, the integration of .fmu models into a PLC run-time is also under development in the case of Beckhoff but suitable alternatives are found. All in all, the data structure works in general and shows its usability in the example. As a result of using the data structure the difficulties in setting up a virtual commissioning can be reduced with the advantage of minimizing time and costs.

VII. OUTLOOK

The next possible steps of this paper are on the one hand further research on the exchange of kinematized CAD assemblies based on a neutral data format and on the other hand further evaluation of the data structure with respect to bigger use cases and additional software for the integration of the behaviour models in the PLC. As part of these larger use cases, compatibility of different versions of the software used can also be further investigate.

REFERENCES

- [1] H. Sangvik, "Digital Twin of 3d Motion Compensated Gangway Use of Unity Game Engine and TwinCAT PLC Control for Hardware-in-the-Loop Simulation," Master's thesis, University of Agder, 2021.
- [2] "Standardized data exchange of CAD models with design intent," *Computer-Aided Design*, vol. 40, no. 7, pp. 760–777, 2008, current State and Future of Product Data Technologies (PDT).
- [3] B. Kim and S. Han, "Integration of history-based parametric translators using the automation APIs," *International Journal of Product Lifecycle Management*, vol. 2, no. 1, p. 18, 2007.
- [4] Y. Kim, H. Lee, M. Safdar, T. A. Jauhar, and S. Han, "Exchange of parametric assembly models based on neutral assembly constraints," *Concurrent Engineering*, vol. 27, no. 4, pp. 285–294, Aug. 2019.
- [5] D. Wang, K. Cao, and J. Wang, "Study on analyzing and remodeling assembly constraints of CAD models from the heterogeneous system," in *International Conference on Computer Vision, Application, and Design (CVAD 2021)*, Z. Zhang, Ed. SPIE, Dec. 2021.
- [6] M. Barnes and E. L. Finch, "COLLADA - Digital Asset Schema Release 1.5.0 Specification." Sony Computer Entertainment Inc., 04 2008.
- [7] Modelica Association Project, *Functional Mock-up Interface Specification, Version 3.0, 2022-05-10*.
- [8] Beckhoff Automation GmbH, *TE1420 — TwinCAT 3 Target for FMI*, product data sheet.
- [9] —, *TE1400 — TwinCAT 3 Target for Simulink*, product data sheet.
- [10] Beckhoff Information System. TE1400—TwinCAT 3 Target for Simulink. Visited on 2022-05-26. [Online]. Available: https://infosys.beckhoff.com/content/1033/te1400_tc3_target_matlab/index.html?id=7328785815492855617



Dominique Geiger is Master student in the Department of Mechatronics at MCI Innsbruck/Austria.

An Educational System for Advanced Control Engineering: Development of a Intelligent Self-Balancing Robot

Christian Gfall, and Yeongmi Kim (supervisor)

Abstract—This thesis develops an advanced control engineering learning system centered on a self-balancing robot. The system includes a speech recognition machine learning model for robot control. The process involves CAD design, fabrication, system modeling, and development of a linear quadratic controller with a Kalman filter observer. The control system is developed using Matlab, while the ESP32 microcontroller is programmed with PlatformIO. The machine learning model, trained with TensorFlow, improves the robot's interaction through keyword recognition and is executed on an Arduino Nano 33 BLE Sense. Instructional scripts produced during development provide practical experience for students. The system enhances learning experiences, provides practical skills, and serves as a prototype for similar systems in other engineering disciplines.

Index Terms—Advanced control engineering, Self-balancing robot, LQR controller, Kalman filter, Machine Learning, Education

I. INTRODUCTION

CONTROL engineering education is constantly evolving and requires the development of innovative teaching methods to enhance student learning. One of the challenges in this field is the complexity of designing and implementing intelligent systems, such as self-balancing robots. While these robots provide a practical way to demonstrate advanced control engineering concepts, they require complicated design and control strategies to maintain balance and respond effectively to commands. Previous work in

this area has focused primarily on the mechanical design and control aspects of these robots. For example, the studies by [1], [2] and [3] have explored various control strategies, including the use of linear quadratic controllers and Kalman filters [4], to maintain a balanced position for these robots. However, integrating a machine learning model for speech recognition to control such a robot is a relatively unexplored area. The hypothesis of this study is that developing a learning system that combines a self-balancing robot with a machine learning model for speech recognition can greatly enhance the learning experience of advanced control engineering students. This system would not only provide students with a practical way to apply control engineering concepts, but also familiarize them with the integration of machine learning into control systems. This paper presents the design and implementation of such a learning system. It describes the development process, including CAD design, fabrication, modeling, linearization, and design of a linear quadratic controller and a linear quadratic estimator, also known as a Kalman filter. The paper also describes the training of a machine learning model for basic instructions and its implementation on an Arduino Nano 33 BLE Sense using the TensorFlow Lite library [5]. The results demonstrate the successful application of this system in a laboratory setting and highlight its potential as a valuable tool for control engineering education.

Y. Kim is with the Department of Medical Technologies, MCI, Innsbruck, Austria, e-mail: yeongmi.kim@mci.edu.

II. METHODS

A. Hardware Design

The focus in designing the robot is on affordability and simplicity of fabrication, with the goal of making the robot accessible for educational purposes. The structure of the robot is mainly made of laser-cut wood, which was chosen for its cost-effectiveness and ease of use. All other components, such as wheels, pulleys or PCB holders, were designed to be manufactured with a 3D printer. The design is optimized to place the center of gravity as far away from the wheel axis as possible to improve the robot's self-balancing ability. This strategic placement of the centre of gravity contributes to the stability of the robot, as it allows the inertia to be largest, making the robot easier to control. The CAD design can be seen in Fig. 1.

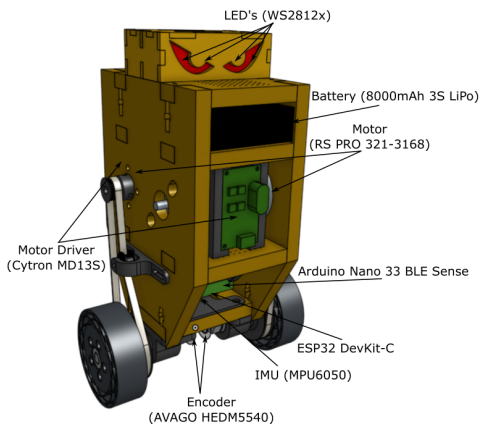


Fig. 1. Rendered picture of the robot, showing the main components.

DC motors, which are widely available and affordable, serve as actuators. The motors and wheel axles are coupled with a belt drive system, a simple and efficient method of power transmission. The belt drive is not only inexpensive, but also easy to assemble and maintain, making it an ideal choice for

this project. An MPU6050 inertia measurement unit (IMU) is installed to measure the robot's position in space and both wheels are equipped with optical quadrature encoders. The IMU has a built in digital motion processor (DMP) is used to gather the angles and angular velocities, rather than the data from the accelerometer and gyroscope. The IMU and encoders are connected to the main microprocessor, a ESP32-Devkit-C V4, which is used to execute the control algorithm. This microcontroller unit (MCU) also controls the Cytron MD13S motordrivers with a pulse-width modulation (PWM) signal for speed and a digital pin for the rotation direction of the motors. An Arduino Nano 33 BLE serves as separate MCU to execute the machine learning (ML) model, because it already provides a built in microphone for keyword spotting. Both microcontrollers are interconnected with a UART-Interface for communication. To provide the user with feedback, four WS2812x LED's are built into the eyes of the robot. These LED's can be controlled via a one-wire bus with the ESP32. The overall electronics layout is illustrated in Fig. 2.

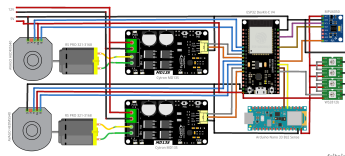


Fig. 2. Overview and connections of the electronics components.

B. Modeling

Modeling is probably the most difficult task when designing a controlled system. Since the purpose of this project is to connect theory with practice, the approach chosen for modeling is as simple and practical as possible, the Newton-Euler approach. This approach allows to understand the dynamics of the robot and how it responds to various forces and

torques. The robot is viewed, as seen in Fig. 3, as an inverted pendulum that can be moved with wheels in the x-direction.

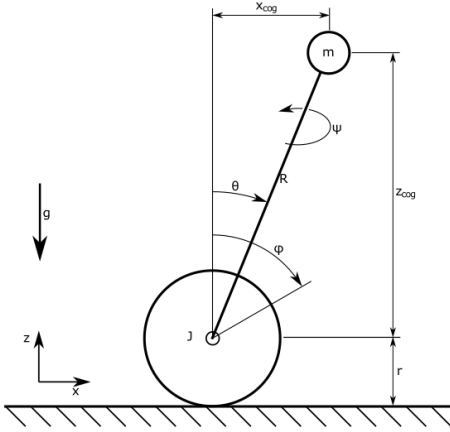


Fig. 3. Simplified sketch that represents the robot as inverted pendulum.

To obtain the equation of motion, the sum of all forces at the fulcrum of the pendulum is needed. Newton's well-known equation of angular motion is used for this purpose:

$$\sum M = m\dot{\omega} = m\ddot{\theta} \quad (1)$$

From this equation, the torque that is generated by gravity (M_g) can be depicted:

$$M_g = mgR \sin \theta \quad (2)$$

In this case R represents the distance from the wheel axles to the center of gravity, m the mass of the robot's body and g the gravitation. Furthermore, also the torque generated from a force in x-direction (M_a), that is generated by the motors can be described as

$$M_a = -maR \cos \theta \quad (3)$$

where the linear acceleration a can be calculated

using angular acceleration $\ddot{\varphi}$ and the wheel-radius r :

$$a = \ddot{\varphi}r \quad (4)$$

Since the robot is a dynamic system, inertial forces M_J coming from the body inertia J must also be taken into account:

$$M_J = -J\ddot{\theta} \quad (5)$$

The overall equation of motion can then be calculated by summing up all the forces, leading to:

$$J\ddot{\theta} = mgR \sin \theta - m\ddot{\varphi}rR \cos \theta \quad (6)$$

It is obvious that this is a nonlinear system. By linearization of Eqn. (6) using Taylor series expansion, the system can be linearized in the range of a working point. The working point is the upright position, i.e. when the robot is in balance. Taking Fig. 3 into account each time dependent variable is set to 0, which leads to the following motion equation:

$$J\ddot{\theta} = mgR\theta - m\ddot{\varphi}rR \quad (7)$$

As input the duty cycle d of the motors is used. Thus a correlation between the control signal of the motors and the angular acceleration $\ddot{\varphi}$ is needed. For this the motor is considered as a first order lag element described by the gain K_M and the time constant τ_M :

$$G_M(s) = \frac{K_M}{\tau_M s + 1} \quad (8)$$

This results in equation 9. Here the duty cycle d is substituted by u for a more general representation.

$$\ddot{\varphi} = \frac{K_M u - \dot{\varphi}}{\tau_M} \quad (9)$$

After including the motor model in the equation of motion, the equation is converted to a state space system. The following variables come into place as states:

$$\vec{x} = [\theta \quad \varphi \quad \dot{\theta} \quad \dot{\varphi}]^T \quad (10)$$

Obviously, $\dot{x}_1 = x_3$ and $\dot{x}_2 = x_4$, and for \dot{x}_4 the equation 9 can be rearranged to:

$$\dot{x}_4 = \ddot{\varphi} = \frac{K_M u - \dot{\varphi}}{\tau_M} \quad (11)$$

Substituting x_4 into x_3 then yields the complete system

$$\dot{\vec{x}} = \mathbf{A}\vec{x} + \mathbf{B}u \quad (12)$$

where the state matrix \mathbf{A} and the input matrix \mathbf{B} are described as

$$\mathbf{A} = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ \frac{mgR}{J} & 0 & 0 & \frac{mrR}{\tau_M J} \\ 0 & 0 & 0 & -\frac{1}{\tau_M} \end{bmatrix} \quad (13)$$

$$\mathbf{B} = \begin{bmatrix} 0 & 0 & -\frac{K_M mrR}{\tau_M J} & \frac{K_M}{\tau_M} \end{bmatrix}^T \quad (14)$$

Since all states can be measured, the output matrix \mathbf{C} is a 4x4 identity matrix, and since there is no entry in the direct transition matrix \mathbf{D} , it is zero.

In the next step the system needs to be discretized, because it is used in a digital system. The discretized state and input matrices can be calculated from the continuous-time matrices using the matrix exponential and integral formulas in Eqn. (16). For a given sampling period T_s , they are typically computed as:

$$\mathbf{A}_d = e^{\mathbf{A}T_s} \quad (15)$$

$$\mathbf{B}_d = \int_0^{T_s} e^{\mathbf{A}\tau} \mathbf{B} d\tau \quad (16)$$

The matrices are not calculated analytically, but with the `c2d()`-command in Matlab. The resulting system can then be used to design an optimal controller, a linear quadratic regulator (LQR) to be specific.

C. Linear Quadratic Regulator

The linear quadratic controller (LQR) is a powerful method in modern control theory used to compute optimal gains for the control of linear systems. The LQR controller is designed based on the state space model in Eqn. (14). The LQR design problem is formulated as an optimization problem, where the objective is to minimize a cost function. The cost function is usually defined as the sum of the squares of the states \vec{x} and the control inputs u , each weighted by a positive definite matrix. As the controller will

be implemented to a microcontroller, the discrete cost function J is given by

$$J = \vec{x}^T(N) \mathbf{S} \vec{x}(N) + \sum_{k=0}^{N-1} \left[\vec{x}^T \mathbf{Q}(k) \vec{x}(k) + \vec{u}^T(k) \mathbf{R} \vec{u}(k) \right] \quad (17)$$

where \mathbf{Q} and \mathbf{R} are the state and control weighting matrices, respectively. The matrices \mathbf{Q} and \mathbf{R} are chosen as a function of the desired performance of the system. The entries in \mathbf{Q} are penalizing large deviations in the states, while the entries in \mathbf{R} penalize large control inputs. \mathbf{R} is only a scalar value for this system, because it has only one input. The solution to the LQR problem is given by the discrete Riccati equation:

$$\mathbf{P}(k) = \mathbf{A}^T \mathbf{P}(k+1) [\mathbf{A} - \mathbf{B} \mathbf{K}(k)] + \mathbf{Q} \quad (18)$$

where \mathbf{K} is the optimal state feedback gain matrix. The LQR controller provides a systematic way to design controllers for linear systems. The controller is optimal in the sense that it minimizes the cost function, which is a trade-off between system performance and control effort. To obtain the best possible performance, the best combination of weights in the \mathbf{Q} and \mathbf{R} matrices must then be tuned empirically. Due to the fact, that a LQR controller is a fullstate controller and the measurements are expected to be noisy, an observer is needed. For this task also an optimal solution, a linear quadratic estimator (LQE), is chosen.

D. Linear Quadratic Estimator

The Kalman Observer, also known as linear quadratic estimator (LQE), is an optimal observer that provides estimates of the internal states of a linear system from noisy measurements. The estimator is designed like a LQR controller, except that the observer gain matrix \mathbf{L} is searched by minimizing a cost function. In the case of the Kalman observer, the \mathbf{Q}_{Obs} matrix represents the measurement noise covariance matrix and \mathbf{R}_{Obs} the process noise covariance matrix. The Kalman filter provides an optimal

estimate of the states in the sense that it minimizes the mean square error of the state estimates. Usually the Kalman filter is working in five steps:

- 1) Prediction of the new error covariance matrix:

$$\mathbf{P}'(k+1) = \mathbf{A}\mathbf{P}(k)\mathbf{A}^T + \mathbf{Q}_{Obs} \quad (19)$$

- 2) Calculation of the new observer gain matrix \mathbf{L} . Here the condition is, that the covariance becomes a minimum:

$$\mathbf{L}(k+1) = \mathbf{P}'(k+1)\mathbf{C}^T \cdot [\mathbf{C}\mathbf{P}'(k+1)\mathbf{C}^T + \mathbf{R}_{Obs}]^{-1} \quad (20)$$

- 3) Prediction of the new states by using the prior estimation and the current input:

$$\hat{\tilde{\mathbf{x}}}'(k+1) = \mathbf{A}\hat{\tilde{\mathbf{x}}}(k) + \mathbf{B}u(k) \quad (21)$$

- 4) Update of the error covariance matrix:

$$\mathbf{P}(k+1) = [\mathbf{I} - \mathbf{L}(k+1)\mathbf{C}]\mathbf{P}'(k+1) \quad (22)$$

- 5) Update of the estimated states by using the predicted states and the estimation error, that is multiplied with the observer gain matrix:

$$\hat{\tilde{\mathbf{x}}}(k+1) = \hat{\tilde{\mathbf{x}}}'(k+1) + \mathbf{L}(k+1)[\hat{y}(k) - \mathbf{C}\hat{\tilde{\mathbf{x}}}(k)] \quad (23)$$

To control the robot's rotation around the vertical axis a proportional-integral-derivative (PID) controller is employed. The PID controller determines the difference between the desired and actual yaw angle ψ , generating a control output. This output adjusts the robot's wheel setpoints, subtracting from the left wheel and adding to the right, enabling yaw angle control through differential rotation. The overall control system is illustrated in Fig. 4, where the measured yaw angle ψ is represented by y_5 .

E. Machine Learning

The machine learning (ML) algorithm for speech recognition is developed using TensorFlow in a Jupyter notebook. An example from an O'Reilly

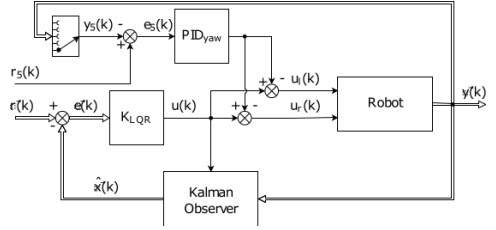


Fig. 4. Scheme of the control system using a LQR controller for balancing, a PID-controller for yaw control and a Kalman Observer.

book[5] is used as the basis for the designed script. The algorithm is trained with audio examples downloaded from the *speech_commands* library[6] of TensorFlow. This library contains 100000 one-second audio samples of random people speaking specified words. Audiosamples with the voice commands *go*, *stop*, *right*, *left*, *forward* and *backward* are used for training. The voice samples are split into a training set, a validation set, and a test set in the ratio 80% : 10% : 10%. The audio files are in WAV format with a sampling rate of 16 kHz. The first step of the process is to perform a short-time Fourier transform (STFT) on the audio samples. The STFT is a mathematical algorithm that divides the audio samples into small time windows and outputs a spectrogram of each window[7]. This can be used to create an image from the individual spectrograms, which is then used to train the ML algorithm. In Fig. 5, the original waveform and the corresponding spectrogram of the word *right* are shown. A Convolutional Neural Network (CNN) is utilized for this model, because it is ideal for image and audio processing tasks. It is crucial to maintain a small memory footprint, hence a model with fewer parameters is chosen. A two-dimensional convolution layer is used due to its lower parameter requirements compared to a dense layer. This Conv2D layer extracts features from the input image, a spectrogram in this case. The values from the STFT are processed by the convolutional layer to highlight features and then passed through a Rectified

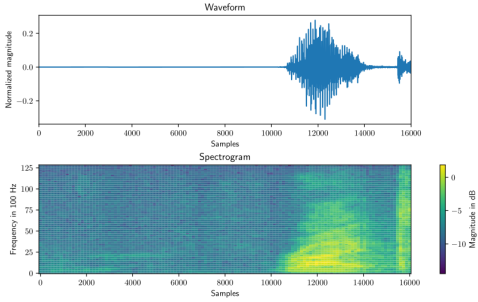


Fig. 5. Waveform and spectrogram of a audio sample containing the word *right*.

Linear Unit (ReLU) activation function for transformation, aiding faster convergence in learning. The model is trained in 15000 steps, with the initial 12000 steps at a learning rate of 0.001. A higher learning rate accelerates learning but makes parameter convergence challenging, so the final 3000 steps are trained at a reduced learning rate of 0.0001. Every 1000 steps during training, a performance test is conducted on a validation dataset to measure the model’s accuracy. This approach allows for performance evaluation and overfitting detection, as overfitting would result in decreased accuracy when tested with the validation dataset. After training, a performance evaluation is conducted using the test dataset.

III. RESULTS

This chapter discusses the results of the different phases of the project, from the design of the robot to the deployment of the machine learning model.

A. Keyword Spotting Performance

The machine learning model’s performance was assessed using accuracy and cross entropy during the training phase. Accuracy measures the ratio of correct predictions to total predictions, while cross entropy is a loss function that quantifies the difference between predicted and actual probability distributions.

A lower cross entropy indicates a closer match between predicted and actual distributions, suggesting a better model. The model was initially trained with six keywords, but the model topology used is not accurate enough to differentiate between the categories. Due to a bug in the TensorFlow-Lite library it is not possible to change the model topology. Therefore *forward* and *backward* are excluded from training. The training was then repeated with the reduced dataset. The accuracy and cross entropy plots showed convergence at the end of training, indicating no overfitting. The model was further tested with the test dataset, that was not used for training, to ensure unbiased evaluation. The test involved 2141 audio samples, contrasting the correct word with the model’s prediction to measure accuracy and create a confusion matrix. The test achieved an overall accuracy of 84.1%. From the confusion matrix in Fig. 6, it can be seen that incorrect classifications tend to happen more frequently, especially when it comes to unknown terms.

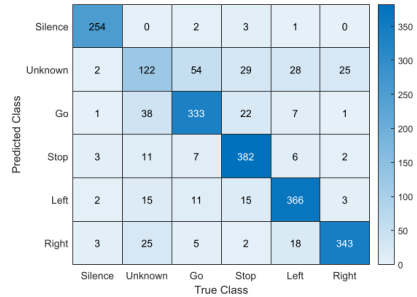


Fig. 6. Confusion matrix of the evaluation run with the test dataset.

B. Control System

After manufacturing the components, the weight of all components was measured and transferred to the CAD model. The values in Tab. I could then be obtained from the CAD model. Motor identification is the final step before the final state space model

TABLE I
VALUES TAKEN FROM CAD DESIGN

Var.	Value	Description
r	Wheel radius	0.0405 m
R	Distance wheel to COG	0.135 75 m
m	Mass of robot chassis	2.848 kg
J	Inertia of the chassis	0.065 07 kg m ²

can be calculated. The time constant τ_M and the motor gain K_M are derived from the step response, using the `tfest()` function in Matlab. For curve-fitting the measurement of a step response of the motor velocity is used. The values $K_M = 1.7262$ and $\tau_M = 0.0704$ were obtained, with a curve-fitting accuracy of 94.736%. The state space model and controller are then calculated using these values and the analytical state space model from Eqn. (14). For discretization using the `c2d()`-function in Matlab a sampling time of $T_s = 0.005$ s is chosen. From this the discrete state equation of the system is obtained:

$$\vec{x}(k+1) = \mathbf{A}_d \vec{x}(k) + \mathbf{B}_d u(k) \quad (24)$$

where the state matrix \mathbf{A}_d and the input matrix \mathbf{B}_d are described as follows:

$$\mathbf{A}_d = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 58.28 & 0 & 0 & 4.10 \\ 0 & 0 & 0 & -17.05 \end{bmatrix} \quad (25)$$

$$\mathbf{B}_d = [0 \quad 0 \quad -2.61 \quad 10.85]^T \quad (26)$$

The LQR controller is designed for the system using the `dlqr()`-function in Matlab, which calculates the optimal gain matrix \mathbf{K} for the controller. After numerous tests, a suitable combination was found that allows the robot to react appropriately to disturbances. The state weighting matrix \mathbf{Q} and control weighting matrix \mathbf{R} gathered from testing as well as the resulting controller gain matrix \mathbf{K} are represented in Eqn. (27), Eqn. (28) and Eqn. (29).

$$\mathbf{Q} = \text{diag}(100 \quad 5 \quad 50 \quad 25) \quad (27)$$

$$\mathbf{R} = 1 \quad (28)$$

$$\mathbf{K} = [-247.84 \quad -1.49 \quad -32.78 \quad -4.32] \quad (29)$$

The largest weight is placed on the pitch angle $\theta = x_1$, because it is most important for the stability of the robot. In addition to the LQR controller, the PID controller for controlling the yaw angle ψ has also been optimized. A combination of a proportional gain of $K_P = 15$, an integral gain of $K_I = 7.5$ and a differential gain of $K_D = 0.1$ was found to be a good compromise between settling time and overshoot.

While testing the controller, the Kalman observer was also tuned. Since the measurement noise covariance matrix \mathbf{Q}_{Obs} and the process noise covariance matrix \mathbf{R}_{Obs} are constant and infinite estimation time is assumed, the observer gain matrix \mathbf{L} can be calculated with the `kalman()`-command in Matlab. The covariance matrices for measurement noise, denoted as \mathbf{Q}_{Obs} , and process noise, represented by \mathbf{R}_{Obs} , which were obtained from testing, along with the derived observer gain matrix \mathbf{L} , are depicted in Eqn. (30), Eqn. (31), and Eqn. (32) respectively.

$$\mathbf{Q}_{\text{Obs}} = 50 \quad (30)$$

$$\mathbf{R}_{\text{Obs}} = \text{diag}(0.05 \quad 0.05 \quad 0.5 \quad 0.5) \quad (31)$$

$$\mathbf{L} = \begin{bmatrix} 0.0225 & 0.0004 & 0.0171 & 0.0007 \\ 0.0006 & 0.0313 & -0.0010 & 0.0132 \\ 0.1679 & 0.0075 & 0.1632 & -0.1428 \\ 0.0199 & 0.0478 & -0.1473 & 0.7107 \end{bmatrix} \quad (32)$$

C. Disturbance Rejection

The robot was subjected to several forward and backward pushes to test its disturbance rejection capabilities. The robot's pitch angle $\theta = y_1$, displacement from the target position $\phi = y_2$, and recovery time were monitored. The system's input effort u to restore equilibrium was also observed. The robot was pushed forward at 6.2 s and 9.5 s, and backward at 14.5 s, as indicated by the three peaks in Fig. 7. The maximum tilt was 0.22 rad or 12.6°. The robot managed to rebalance and return to its

starting position within approximately 2 s after each disturbance. The motors operated at a maximum duty cycle of 19 %, demonstrating efficient performance. However, the motor position y_2 showed a steady-state offset of about 2 rad, meaning the robot was approximately 0.08 m ahead or behind its intended position. Despite the presence of system noise, as seen between $t = 4$ s and $t = 6$ s, the Kalman observer ensured that this noise has minimal impact on the robot's performance.

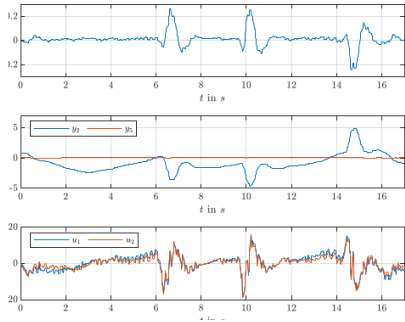


Fig. 7. Rejection of pulse-like disturbances. The robot was pushed backward at $t = 6.2$ s and $t = 9.5$ s and forward at $t = 14.5$ s.

D. Dynamic Behaviour

The robot is designed to respond to voice commands, specifically the keywords *go* and *stop* to initiate forward or backward movement, and *left* and *right* for rotation around the Z-axis of the robot. Two experiments were conducted to analyze the robot's behavior. In the first experiment, the robot's forward and backward driving behavior was examined. The robot was given a *go* command at $t = 8$ s and a *stop* command at $t = 20$ s, causing it to move forward or backward by one meter. This corresponds to a position setpoint of $r_2 = 24.69$ rad. As can be seen in Fig. 8, the target position was reached approximately $t = 9$ s after both commands were issued. The robot's acceleration and braking motions

could be observed from the pitch angle y_1 and the position y_2 . To accelerate, the robot leans forward by initially accelerating in the opposite direction, then applies acceleration in the intended direction, returning to an upright position and gaining speed. This process involves a maximum tilt of 6.3° . In the

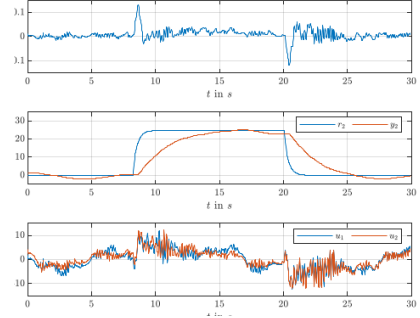


Fig. 8. Response to a *go*-command at $t = 8$ s and a *stop*-command $t = 20$ s.

second experiment, the robot responds to left and right commands by rotating 90° (1.57 rad) around its Z-axis. Fig. 9 shows that the target angle is achieved within $t = 2$ s for all voice commands. The amplitudes of duty cycles u_1 and u_2 illustrate the PID controller's differential action to offset the yaw angle.

E. Developed Learning Methods for Students

The whole Matlab code was written in a Matlab Live Script that serves as an interactive tool for students to learn controller design and modeling. The script provides a comprehensive step-by-step guide, allowing students to interact with the code and observe the results in real time. The Live Script also includes visualizations such as plots and diagrams that facilitate understanding of the concepts used. By working with the live script, students can modify the code, experiment with different parameters, and observe the effects on the controller's performance. This interactive learning approach encourages exploration

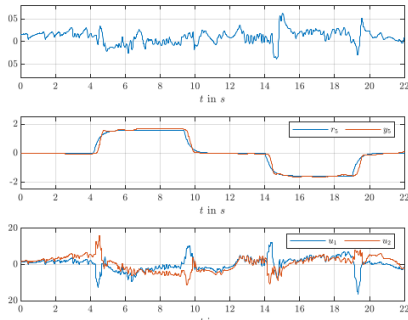


Fig. 9. Response to a *left*-command at $t = 4.5$ s and $t = 19$ s and a *right*-command $t = 9.5$ s and $t = 14$ s.

and deepens understanding, resulting in an engaging and effective learning experience. The machine learning component of this project was developed in a Jupyter script. Jupyter notebooks are known for their ability to combine code, visualizations, and narrative text, making them an ideal tool for teaching complex topics such as machine learning. The Jupyter script developed for this project guides students through the process of implementing a keyword spotting algorithm. It includes detailed explanations of the code and theoretical concepts behind each step.

IV. CONCLUSION

The development of this advanced control engineering learning system has enhanced students' learning and provided practical skills. However, the approach has limitations. The robot design was based on specific parameters that could affect performance if changed. The machine learning model was trained on a specific data set, so its performance may vary in real world conditions. Going forward, optimizing the robot design and expanding the machine learning model with more commands are promising research directions. One main finding was the robustness against disturbances of the controller, emphasizing the performance that is achievable with a LQR controller, even if the system is highly unstable.

REFERENCES

- [1] A. A. Bature, S. Buyamin, M. N. Ahmad, and M. Muhammad, "A comparison of controllers for balancing two wheeled inverted pendulum robot," *International Journal of Mechanical & Mechatronics Engineering*, vol. 14, no. 3, pp. 62–68, 2014.
- [2] F. Sun, Z. Yu, and H. Yang, "A design for two-wheeled self-balancing robot based on kalman filter and lqr," in *2014 International Conference on Mechatronics and Control (ICMC)*. IEEE, 2014, pp. 612–616.
- [3] M. Önkol and C. Kasnakoğlu, "Adaptive model predictive control of a two-wheeled robot manipulator with varying mass," *Measurement and Control*, vol. 51, no. 1-2, pp. 38–56, 2018.
- [4] F. Auger, M. Hilaret, J. M. Guerrero, E. Monmasson, T. Orłowska-Kowalska, and S. Katsura, "Industrial applications of the kalman filter: A review," *IEEE Transactions on Industrial Electronics*, vol. 60, no. 12, pp. 5458–5471, 2013.
- [5] P. Warden and D. Situnayake, *Tinyml: Machine learning with tensorflow lite on arduino and ultra-low-power microcontrollers*. O'Reilly Media, 2019.
- [6] P. Warden, "Speech Commands: A Dataset for Limited-Vocabulary Speech Recognition," *ArXiv e-prints*, Apr. 2018. [Online]. Available: <https://arxiv.org/abs/1804.03209>
- [7] M. Portnoff, "Time-frequency representation of digital signals and systems based on short-time fourier analysis," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 28, no. 1, pp. 55–69, 1980.

Christian Gfall is a student of the Mechatronics & Smart Technologies program at the MCI Innsbruck/Austria. The topics of his work are largely related to control engineering and robotics. In his current master's thesis, he is also transferring this knowledge to the field of machine learning.



Experimental determination of the material parameters of an alpine ski

Gregor Hofmeister, and Franz-Josef Falkner (supervisor)

Abstract—Prototyping an alpine ski is time consuming and expensive. This problem can be reduced if the bending and torsional stiffness of new materials can be predicted from previous simulations. The characterisation of the materials used is a major problem for simulation. The results of an experimental modal analysis and 3 point bending test of a wooden core, were compared to simulations with *Ansys Workbench*. From this, the material parameters for bending stiffness are determined and a characterisation can be performed. This analysis is integrated into a developed FE program, called *FEMski*, which can confirm the bending stiffness of the materials and can be used for future component calculations. The material characterisation by modal analysis show good results, allowing this structural analysis to be used alongside a 3-point bending test. The *FEMski* software determines a bending stiffness that is about 2.3% higher than the simulation with *Ansys Workbench*. With this result, *FEMski* can be used to simulate the bending stiffness. The main objective, a simple material characterisation as well as its integration into the bending stiffness calculation of the material of an alpine ski, has been achieved with this work. The simple characterisation and bending stiffness calculation can save time and money in the development of prototypes. This approach is to be pursued further in order to make even better statements about the bending lines and torsional stiffness of an alpine ski.

Index Terms—Alpine ski, FE simulation, material characterization, modal analysis, bending stiffness

I. INTRODUCTION

FOR structural and modal simulation of alpine ski FE methods already have been used. Especially

G. Hofmeister studies at Management Center Innsbruck, e-mail: g.hofmeister@mc4me.at.

Dr. techn. F. Falkner, Lecturer at Management Center Innsbruck, e-mail: Franz-Josef.Falkner@mc1.edu.

to investigate the interaction between athlete, ski and snow by implementing snow - ski boundary conditions, different FE simulations are used [1], [2], [3], [4]. The simulation of the actual turn radii have been developed in dependence of snow hardness, skier's body weight, ski mechanics and edging angle [2], [5], [6]. There are still limitations in modeling the interaction between ski and snow. Numerical instabilities due to changing boundary conditions are one of the reasons why some authors consider the application of their simulations problematic [2]. Mössner et al. were able to overcome the limitation of the simplified snow reaction force and the irregularity in the interface between ski and snow. Despite all the simulation progress, it is not yet a standard in the ski industry to model the ski-snow interaction [7]. Increased computational power and today's ease of use of FE software seem to make structural simulations of alpine skis feasible even for small producers. There are published methods of modeling ski mechanics [8],[9]. However, to the author's knowledge, these methods remain not a standard in commercial software application. In addition, significant discrepancies between practical and numerical results of mechanical ski parameters of up to 20% in torsion and 15% in bending have been found in research [2]. For these reasons, valid simulation of ski structure behavior remains a challenge. With this scientific work, methods for a better material characterisation will be developed. Additionally, a FE-simulation for the bending stiffness determination of materials will be used. A new software, called *FEMski*, is used with which statements about the bending stiffness of the product can be made, as well as the bending line

can be represented. The results will be compared with experimental results in order to be able to make possible statements about the usability. Furthermore, the possible limits and the required effort will be shown.

II. METHODS

A. Model

An existing all-mountain ski with unknown exact material parameters is the basement for this work. A CAD model of the geometry is analysed from provided 3D files by the ski manufacturer. The ski is modelled over its full length included with shovel and tail. The detailed sandwiched construction is modelled by all the different solid bodies with bonded contacts which are consisting of wood core, resin interlayers, sidewalls, upper and lower surface and edges. The *Original+* company guarantees to change the characteristics of the ski by using different materials. Depending on the type of wood used, the damping characteristics can be influenced. The torsional properties are influenced by different placement of the *titanal* inserts and different thicknesses and fiber arrangements of the fiberglass, carbon and flat inserts. The deflection of the ski due to pressure is influenced by the shape of the wood core and the type of wood and different laminates of *titanal*, fiberglass, carbon or flax [10]. The 3D model can be seen in detail on Fig. 1.

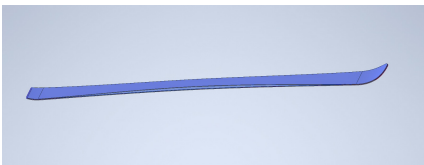


Fig. 1. 3D-Model of the ski

This ski model is examined for vibrations and damping behaviour with a experimental modal analysis. Furthermore, the woodcore will be investigated with the experimental modal analysis as well with

an 3-point bending test. With the corresponding simulation, the material parameter will be determined. The last part contains the calculation of the bending stiffness of the timber core with *FEMski* and the comparison to the other simulations.

The wood cores examined in this and the following chapters vary in their thickness distribution along their length. For example, core *B* has a thickness of 10.96 mm at the thickest point after milling, while core *S* has a thickness of 11.96 mm at the thickest point. Among other things, the influence of the thickness distribution on the material properties is investigated.

B. Laboratory Tests

For comparison with the simulations a 3-point bending test and a experimental modal analysis is performed. Each laboratory test will be explained in the following shortly.

1) *3-point bending test*: For the 3-point bending test the object to be examined is mounted freely movable on two round steels at the front and rear, see Fig. 2. The force is applied via an electromechanical linear drive in predefined deflection steps until a maximum value is reached. The force reaction is measured at the position where the deflection is applied. With each deflection step the force difference is measured. In addition to the bending characteristic value defined according to DIN ISO 5902, the bending test also provides the bending line (shape of the loaded ski) in the loaded and unloaded state, from which the bending stiffness distribution is derived.

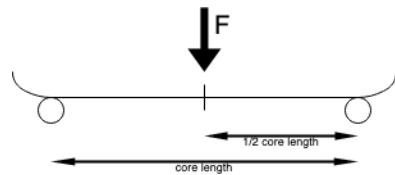


Fig. 2. Setup scheme 3-point bending test

The practical 3-point bending test shown in figure

3 is based on the ISO guidelines 12135, 12737 and 178 [11], [12], [13].

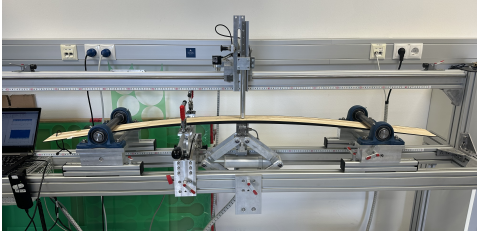


Fig. 3. Photo of a practical 3-point bending test

The distance between the two bearing positions were changed in three steps. This distance change was applied to get a wide range for measurement results and to generate different bending lines.

2) *Experimental Modal analysis*: The wood core, including the attached sidewalls, was hung on a rope through a hole in the upper part of the core. This way, the ski could vibrate freely and the surroundings had the least influence on it. Two accelerometers of the type *DeltaTron Type 4517* from *Brüel & Kjær* were attached 5 cm underneath the hole to measure the movement of the wood core. The wood core and an overview of the measurement setup is shown in Figure 4. The position of the accelerometers is marked with red dots. To generate a vibration an impact hammer of type 8206-01 from *Brüel & Kjær* was used.



Fig. 4. Photo of the easurement setup of modal analysis of wood core, red points = position of the accelerometers

For the experimental modal analysis on an alpine ski the ski was suspended by two ropes. This serves for free oscillation without predefined boundary conditions. Again, two accelerometers of type *DeltaTron Type 4517* from *Brüel & Kjær* were attached. First the accelerometer was positioned at the green dots

in the picture 5 and after one measurement the accelerometer was attached to position 2 (blue points). The ski was always hit with the same impact hammer at the same point. The red point in Figure 5 is in between the two positions of the accelerometers and this is the area where the impact hammer hit. Additionally, it needs to be mentioned that the goal was to generate the impact slightly off-centered to get more torsional vibrations. An impact in the middle area of the ski would just generate bending vibrations, which is the goal to measure, but torsional vibrations are also of great interest.

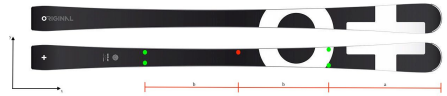


Fig. 5. Positions of the accelerometer and impact hammer. Red point = point where impact hammer was hit, green points = position 1 of accelerometer, blue points = position 2 of accelerometer. $a = 40$ cm, $b = 44$ cm

C. Simulation

1) *Simulation of the 3-point bending test*: The simulation is based on the experimental 3-point bending test. The three different bearing positions were attached to the two different wood cores using *Fusion 360*. The models were then imported into *Ansys Workbench* and a "Static Structural" block was used to integrate the execution of the test. For each bearing position the simulation was carried out for both woodcores. The difference was to change the material parameter of the Young's modulus. Five simulations with different Young's moduli were performed. The test setups and deflections remained the same, only the engineering data was changed. The different Young's moduli are listed in Table I.

TABLE I
FIVE DIFFERENT YOUNG'S MODULI USED FOR
SIMULATION 3-POINT BENDING TEST

Young's Modulus in MPa				
9400	9425	9450	9475	9500

The same deflection as in the experimental 3-point bending test for each bearing position and wood core in mm are inserted as a so-called "Remote Displacement". The aim of the simulation is to measure the force acting in the z-direction, which acts during deflection. This data is then evaluated using *Microsoft Excel*, as in Chapter II-B1. Again, the force difference per deflection step is of interest. These differences are evaluated in a graph and the spring constant k is determined with the aid of the slope of the regression line. This constant is then determined for different Young's moduli and compared with the k from the practical 3-point bending tests.

2) *Determination of the bending stiffness of a wood core from an alpine ski:* As a basis for the simulation, various data must be processed in order to successfully calculate the bending stiffness. A *Microsoft Excel* file containing a list of material parameters with predefined names is created. The Young's and shear modulus are stored in this list. The imported model will be sliced in predefined steps. For the present simulation, a cross section was performed every 30 mm. The first and last 20 mm are cut off, to be sure to generate big enough cross sections. After performing the slicing, the individual cuts can be viewed in detail. After performing these steps, a Finite Element Method can be started. The bending stiffnesses per cross-section are calculated. In addition, the torsional stiffness of each cross-section is determined. This can be used for further research work. After a successful calculation, the bending stiffness is displayed in a graph via the intersection points. The generated results of the bending stiffness are then used to calculate the bending line in order to calculate the deflection of the imported 3D model at the respective intersection points. For this purpose, the list of the results is imported into a *Matlab* script. First of all, the coordinate systems of the software and the script need to be merged together. Additionally the length l_0 between the bearing positions needs to be inserted. The script excerpt determines the bending line based on a boundary value problem. A integration of a system of differential equations is

applied. The inputs are the bending stiffness as well as the torque progression. Finally, the evaluation of the differential equation calculates the bending line w .

3) *Simulation of the Modal analysis:* The simulation of the modal analysis is carried out using the software *Ansys, Inc.* After importing, the material parameters are assigned to the correct material in the *Ansys Mechanical* interface. The mesh is then generated. The mesh has an element size of 0.03 m, chosen randomly to obtain valid results but also to reduce the computation time. A so-called "free-free" modal analysis is implemented. This means that there are no boundary conditions and the structure is free of any loads [14]. Theoretically, the first six eigenfrequencies of the "free-free" modal analysis must be 0. By solving the analysis numerically with *Ansys Mechanical*, the first three eigenfrequencies are 0, eigenfrequencies four to six are not 0, but very close. This can be seen in Figure 6. The first three eigenfrequencies with a bending eigenmode and the first two eigenfrequencies with a torsional eigenmode are considered for further investigation.

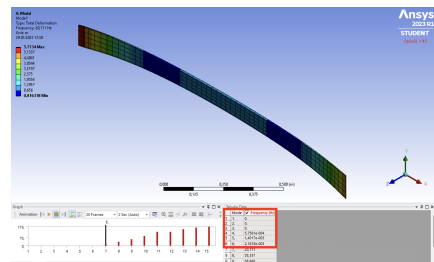


Fig. 6. Implemented "free-free" modal analysis with *Ansys, Inc.* red marked = first 6 eigenfrequencies are close to 0.0 Hz

To fit the simulated frequencies with the measured frequencies out of chapter II-B2 an optimization is introduced. This optimization calculates the target frequencies of the bending and torsional eigenmodes. The Young's modulus for the bending eigenmodes and the shear modulus for the torsional eigenmodes

and the corresponding eigenfrequencies are parameterised. The target frequencies are implemented before the optimization starts. The optimization leads to a variation of simulations of the modal analysis with a varying Young's modulus in case of the bending eigenmodes to achieve the target eigenfrequencies. The same procedure is done for simulating the torsional eigenmodes and the corresponding eigenfrequencies.

III. RESULTS

A. Comparison of modal analysis

First, the measured eigenfrequencies and the corresponding types of eigenmode are shown in Table II.

TABLE II
MEASURED EIGENFREQUENCIES WITH BENDING AND TORSIONAL EIGENMODES

Eigenfrequencies with bending eigenmodes in Hz	
1. eigenfrequency	20, 1
2. eigenfrequency	35, 5
3. eigenfrequency	60, 0
Eigenfrequencies with torsional eigenmodes in Hz	
1. eigenfrequency	70, 0
2. eigenfrequency	87, 2

In comparison to the eigenfrequencies with bending eigenmode, a simulation of the modal analysis and an optimization were done. Two potential candidates for the Young's modulus with associated eigenfrequencies are listed in Table III. The reduction or increase of each eigenfrequency in comparison to the measured eigenfrequencies are listed as well.

TABLE III
OPTIMISED YOUNG'S MODULUS WITH CORRESPONDING EIGENFREQUENCIES WITH DIFFERENCE TO TARGET EIGENFREQUENCIES

First candidate Young's modulus: 9509.5 MPa	
1. Eigenfrequency: 20.157 Hz	+0.2836%
2. Eigenfrequency: 35.434 Hz	-0.1859%
3. Eigenfrequency: 58.849 Hz	-1.9183%
Second candidate Young's modulus: 9509.1 MPa	
1. Eigenfrequency: 20.156 Hz	+0.2786%
2. Eigenfrequency: 35.434 Hz	-0.1859%
3. Eigenfrequency: 58.848 Hz	-1.92%

In particular, the first two simulated eigenfrequencies correspond with the corresponding measured eigenfrequencies with a difference of less than 0.3%. The third simulated eigenfrequency has a difference of less than 2%. Even this difference is still a good approximation to the target frequency. The bigger difference of the third eigenfrequency can be related to the fact that carrying out the practical modal analysis on the wood core was very difficult due to the light weight of the core because activating the vibrations through the impact hammer can create double vibrations. For further investigations during this project the first candidate for the Young's modulus of 9509.5 MPa was used. The simulation and the subsequent optimization of the eigenfrequencies with a torsional eigenmode generated the following results:

TABLE IV
OPTIMISED SHEAR MODULUS WITH CORRESPONDING EIGENFREQUENCIES WITH DIFFERENCE TO TARGET EIGENFREQUENCIES

First candidate shear modulus: 1067.3 MPa	
1. Eigenfrequency: 71.187 Hz	+1.6957%
2. Eigenfrequency: 84.754 Hz	-2.8051%
Second candidate shear modulus: 1066.7 MPa	
1. Eigenfrequency: 71.167 Hz	+1.6671%
2. Eigenfrequency: 84.730 Hz	-2.8326%

Compared to the two eigenfrequencies with torsional eigenmode from Table II, the simulated eigenfrequencies have a difference of less than 2.9%.

These results are very close to the target frequencies. The larger difference compared to the previous optimisation can again be explained by the lightness of the wooden core. The first optimised shear modulus was selected for further investigation with a value of 1067.3 MPa.

B. Experimental modal analysis of an alpine ski

First of all the results of the ski with the slightly bigger woodcore *S* is presented. By comparing the eigenfrequencies just for the first ski between the two measuring positions, it clearly can be stated that the results are close to each other. Only slight differences of under 0.3% between the first natural frequencies of the respective ski can be seen here. The results of the damping factors at the respective natural frequencies are similarly close to each other. The comparison of the second ski the highest difference between the two positions is under 0.08%. These nearly identical results with the corresponding damping factors at both measurement positions for the respective ski suggest that the results of an experimental modal analysis provide the same results regardless of the measurement position. The damping factor of below 1 means that the whole ski as a system is underdamped. An underdamped system, in this case the ski, oscillates with a low amplitude but for a longer time. A closer look at the results of the second ski with wood core *B* shows that the measurement results are very similar to those of the other ski. One big difference can be seen in the slightly smaller eigenfrequencies at both measuring positions. These differences in natural frequency can be attributed to the lower weight of the wood core of just under 20 g, since the natural frequency is calculated as follows and depends on the mass of the system:

$$\omega_n = \sqrt{\frac{k}{m}} \quad (1)$$

C. Results of the experimental 3-point bending test

Comparing the three bearing positions of the 3-point bending test of the wood core *B* a clear dif-

ference in the force required to increase the deformations can be seen. From position 1 to position 3 the needed force drops remarkably, for example for the first 10 mm of deflection from 57.5 N at position 1 to 41.7 N at position 2 and down to 32.2 N at the final position. Similar for the maximum displacement of 45 mm, the needed force of 140.3 N at position 1 drops to 102.9 N at position 2 and finally to 76 N at position 3. At all three measurement positions overshoots can be seen. Which underlines a potential sensor inertia, but as already mentioned a further investigation is needed in this case. Taking a closer look at the smoothed curves of all measurement positions of wood core *S*, it shows a similarity in the force drop from positions 1 to position 3. A detailed comparison of the different wood cores at measurement position 2 shows that the needed force to displace wood core *B* for 10 mm at bearing position 1 has a value of 41.7 N. In comparison to this for the same displacement it needs a force of 45.6 N. This result can be attributed to the thickness distribution of the wood cores. The investigation of the spring constant of each wood core presents a read value for *k* is 2.3096 for wood core *B* and 2.6383 for wood core *S*. This slightly higher value for wood core *S* can again be justified by the greater thickness distribution of the wood core. This higher stiffness of the wood core *S* is confirmed in the other bearing positions. A higher value for the spring constant is determined for all measurements.

D. Results of the simulation of the 3-point bending test

By comparing the measured results for the spring constant with the simulated ones at bearing position 2 and 3, a correlation between the constants can be seen the higher the Young's modulus gets. At bearing position 2 and a Young's modulus of 9500 MPa, a difference of 2.24% occurs and at bearing position 3 and the same Young's modulus the difference result is 3.82%. This means to correlate the spring constants at these two bearing positions a much higher Young's modulus would occur compared to

the Young's modulus determined by modal analysis. However, this contradicts the general basis that the dynamic modulus of elasticity is higher in value than the static modulus of elasticity [15]. All in all, however, all constants lie within a difference range of 5%, so that the differences can be regarded as very small. The simulation of the wood core *S* results in a larger difference to the measured values in the practical test. These higher values for the spring constant *k* indicate a stiffer model in the simulation. However, a deviation of more than 13% is due to a possible error. After further simulations, the results for the wooden core *S* were confirmed. The wood core was then examined for the experimental 3-point bending test. When the thickness distribution was measured again, it was found that in the large thickness range, the differences with the simulation model were up to 15 tenths. The measured wood core has a slightly lower thickness distribution in the centre than the corresponding 3D model. This could be a first explanation for the large differences in the results between the simulation of the 3-point bending test and the experimental investigation. However, further analysis would need to be carried out to provide a concrete explanation for these differences. Figures 7 & 8 showing the bending lines of the respective simulations of the wooden core *B* at bearing position 1. A Young's modulus of 9475 MPa was used to calculate the bending line using *FEMski*. It is easy to see that both curves form a uniform ellipse. This means that there are no irregularities in the calculation, especially for the bending line via *FEMski*. The maximum deflection there is 5.1223 mm, which means that the deflection is 2.39% greater.

Comparing three bending lines of a wooden core determined by *FEMski* with the same Young's modulus of elasticity at three different bearing positions, the difference on the x-axis is striking. As the distance between the bearings changes, the maximum deflections become smaller. When looking at the maximum deflection of the wood cores determined by *FEMski* at different Young's modulus and changed bearing positions with the maximum deflections in

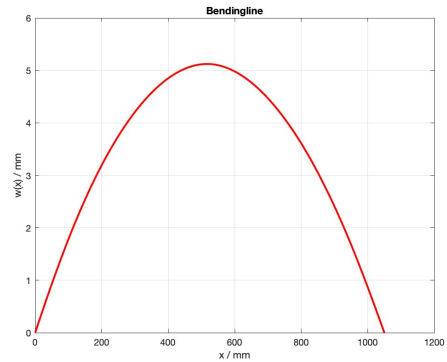


Fig. 7. Bending line via *FEMski* - max. deflection = 5.1223 mm, Young's modulus = 9475 MPa

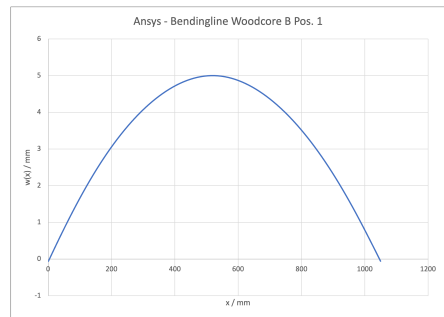


Fig. 8. Bending line via *Ansys Workbench* - max. deflection = 5 mm

the simulation by *Ansys Workbench*, different points stand out. First of all, the differences lie in a maximum of 3.55%, which leads to a good result. Upon closer inspection, the deflections of *FEMski* are all larger than the comparative deflections of *Ansys Workbench*. This could be due to the more powerful calculation of *Ansys* and the number of intersections of *FEMski*. Further investigations would have to determine what influence an increase in the number of cut surfaces in *FEMski* has on the maximum deflection. The results of the bending lines

of woodcore B at bearing position 2 are particularly close to the comparative deflection of 5 mm. All in all the differences range from 0.62% to 3.55%. More precisely, most of the values are in the range of 2-3%. The calculated mean value of 2.37% leads to the assumption that the calculation of the bending stiffness and the resulting bending line by *FEMski* is 2.37% higher than the simulation of the bending line by *Ansys Workbench*.

IV. CONCLUSION

The results of the modal analysis simulation using *Ansys Workbench* gave plausible results for the Young's modulus of the wood core, especially when compared to the 3-point bending test. This shows that modal analysis can be used to determine the parameters of a material. This also makes it easier to determine other material parameters such as shear modulus. However, further research would need to be carried out to confirm this thesis. Modal analysis can also be used to determine the damping properties of complex components. In addition, the possibilities for bending stiffness calculations were investigated using the software program *FEMski*. The results obtained were carried out with 3-point bending simulations by *Ansys Workbench*. Similar results, only slightly higher values, were received of the bending stiffness by *FEMski*. However, as these differences remain within a scientifically acceptable range, the assumption is made that the *FEMski* software program can be used to calculate the bending stiffness. The program scores points for its simple handling and fast calculation times. In addition, the program could also be used to carry out investigations on torsional stiffness.

REFERENCES

- [1] P. A. Federolf, "Finite element simulation of a carving snow ski," Ph.D. dissertation, ETH Zürich, Zürich.
- [2] P. Federolf, M. Roos, A. Lüthi, and et al, "Finite element simulation of the ski-snow interaction of an alpine ski in a carved turn," *The engineering of sport*, vol. 12, pp. 123–133, 2010.
- [3] J. Truong, "Evaluating the directional stability of alpine skis through the simulation of ski deformation during a steady-state turn," presented at the 12th Conference of the International Sports Engineering Association, Brisbane, Queensland, Australia International Sports Engineering Association, 2018.
- [4] Y. Hirano and N. Tada, "Numerical simulation of a turning alpine ski during recreational skiing," *Med Sci Sports Exerc*, vol. 28, pp. 1209–1213, 1996.
- [5] P. Federolf, A. Lüthi, M. Roos, and A. et al, "Parameter study using a finite element simulation of a carving alpine ski to investigate the turn radius and its dependence on edging angle, load, and snow properties," *The engineering of sport*, vol. 12, pp. 135–141, 2010.
- [6] M. Mössner, D. Heinrich, K. Schindelwig, and et al., "Modeling of the ski-snow contact for a carved turn," *The engineering of sport*, vol. 6, pp. 195–200, 2006.
- [7] M. Mössner, D. Heinrich, K. Schindelwig, and et al, "Modeling the ski-snow contact in skiing turns using a hypoplastic vs an elastic force-penetration relation," *Scand J Med Sci Sports*.
- [8] A. A. Nordt, G. S. Springer, and L. P. Kollár, "Computing the mechanical properties of alpine skis," *The engineering of sport*, vol. 2, pp. 65–84, 1999.
- [9] —, "Simulation of a turn on alpine skis," *The engineering of sport*, vol. 2, pp. 181–199, 1999.
- [10] (2022) Original+ mod.8 all mountain ski. [Online]. Available: <https://www.original.plus/produkte/pisten-ski.html>
- [11] ISO Central Secretary, "Metallic materials — unified method of test for the determination of quasistatic fracture toughness," International Organization for Standardization, Geneva, CH, Standard ISO 12135:2021, 2021. [Online]. Available: <https://www.iso.org/standard/78208.html>
- [12] —, "Metallic materials — determination of plane-strain fracture toughness," International Organization for Standardization, Geneva, CH, Standard ISO 12737:2010, 2010. [Online]. Available: <https://www.iso.org/standard/55919.html>
- [13] —, "Plastics — determination of flexural properties," International Organization for Standardization, Geneva, CH, Standard ISO 178:2019, 2019. [Online]. Available: <https://www.iso.org/standard/70513.html>
- [14] I. Ansys, "Free-Free Modal and Pre-stressed Modal Analysis," Skriptum, 2020, Ansys, Inc.
- [15] Z. W. et al., "Measurement of dynamic modulus of elasticity and damping ratio of wood-base composites using the cantilever beam vibration technique," *Construction and Building Materials*, vol. 28.

Development of a Vacuum-tight Exchange Mechanism for Mass Spectrometry

Johannes Höglauer and Sebastian Repetzki (supervisor)

Abstract—The aim of the Master Thesis is to develop a Vacuum-tight Exchange Mechanism for Mass Spectrometers. The project is done at Ionicon Analytik GmbH, the producer of mass spectrometers.

The maintenance of these machines requires the flooding of the vacuum inside the systems, which causes very time consuming pumping processes after a restart. Goal of the thesis is to reduce the time needed, to maintain or exchange the reaction chamber, a key component, of the machine. The state of knowledge, concerning the technical implementation of mass spectrometry, gasket materials and construction methods is analysed by literature research.

Further, the thesis focuses on the practical prototype development of an exchange mechanism. A concept for the gasket and for the mechanics is described. The chosen concept is a baseplate holding two reaction chambers, that slides over a PTFE-gasket. An electric drive carries out the kinematics. Using pressure cells, the vacuum-tightness of the whole system is evaluated by different leakage tests. The results of the leakage tests reveal, that the designed gasket is not yet as tight as the standard configuration - but still in a sufficient range. Furthermore, the gasket is subjected to a durability test to assess, whether the gasket withstands long-term operation. This test shows, that the gasket withstands a minimum of 100 cycles. Concerning the construction of the mechanical components, the manufactured system offers a functional and reliable prototype, providing equal measuring results as the standard configuration. Requirements, which are not met steadily, are the repeatability of the signal intensity, and the consistency concerning the switch-on behaviour of the ion source after the exchange process. Further research is necessary to investigate, whether instrument instabilities or the designed concept cause these issues. Regarding the main purpose of the thesis, results show, that exchange between two reaction chambers is possible in 12 min.

J. Höglauer study at the Department of Mechatronics, MCI, Innsbruck, Austria, e-mail: hj8579@mci4me.at

The automated process of out- and in-coupling of one reaction chamber takes 3 min and 30 s, from measurement to measurement. Concluding, the thesis shows the huge potential of the implementation by producing valid measurement data and successfully carrying out a fast exchange process of the reaction chambers.

Index Terms—Mass Spectrometry, vacuum-tight, Exchange Mechanism, Gasket, PTFE-Seal

I. INTRODUCTION

MASS spectrometry (MS) is used to determine the mass-to-charge ratio. Ionicon Analytik uses the proton-transfer-reaction (PTR-MS) for ionisation [1]. With this method, organic compounds can be ionised in the gaseous state. The ions then undergo mass analysis in a Time-of-Flight (TOF) mass spectrometer. Figure 1 shows the schematic of a PTR-MS with a Time-of-Flight analyzer.

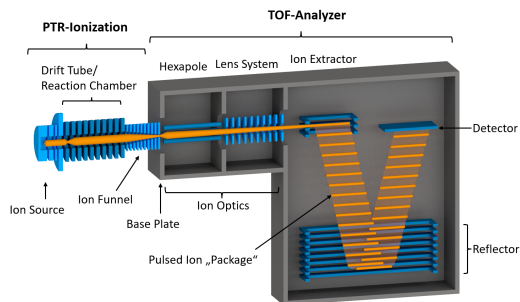


Fig. 1. Proton-Transfer-Reaction Mass Spectrometry instrument (PTR-MS) with a Time-of-Flight analyzer. The orange beam is the trajectory of the ionised molecules.

The reaction chamber (PTR Drift Tube) is firmly screwed onto the Time-of-Flight mass spectrometer. Goal of the thesis is the development of a vacuum-tight exchange mechanism, which allows to switch between two reaction chambers without flooding the vacuum inside the instrument. Shut down, flooding and long pumping after restart is not necessary anymore.

II. METHODS & REALIZATION

A. Mechanical Design

The developed mechanism contains a baseplate, with two reaction chambers mounted, that slides on a PTFE gasket. A electrical actuator moves the baseplate from one position to the other. The CAD design is depicted in Figure 2.

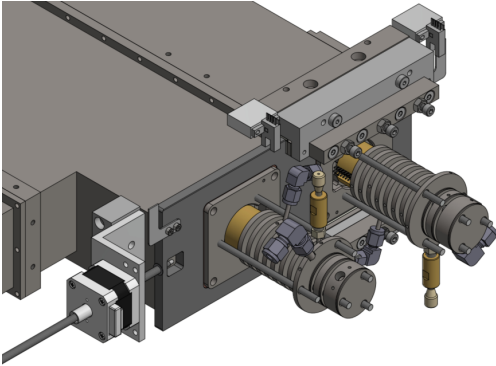


Fig. 2. Mechanical design as CAD view.

The sliding baseplate, is pressed down onto the gasket with six compression springs, allowing long-term, minimal material settlement of the gasket. The contact pressure is calculated as follows:

$$F_{total} = F_{vacuum} + F_{clamp} \quad (1)$$

The ambient pressure also causes a force on the plate with the area A . Δp is assumed as 1 bar. The force holds as:

$$F_{vacuum} = A\Delta p \quad (2)$$

The sliding guides are pressed down by the springs with a known spring constant k and compressed with the spring deflection Δs , resulting in the clamping force

$$F_{clamp} = 6k\Delta s \quad (3)$$

The vacuum force is $F_{vacuum} = 342 \text{ N}$. The spring constant k is chosen as 37.27 N/mm , allowing a maximal clamping force F_{clamp} range of $0-1270.8 \text{ N}$. According to the results of a test using varied clamping forces, the additional force of the springs is chosen as $F_{clamp} = 280 \text{ N}$. The sum of the forces is $F_{total} = 622 \text{ N}$ and is crucial for the dimensioning of the electrical drive. The first prototype is made of aluminium.

B. Actuation

The feed force F_{feed} of the electrical drive is calculated by:

$$F_{feed} = \mu F_{total} \quad (4)$$

With the friction coefficient μ of the material pairing PTFE-Steel and the total force F_{total} , calculated in Equation 1.

The positioning tolerance T_a of the drive system is calculated as the sum of the single tolerances t_i of each component, holding a "worst case" tolerance of [2]:

$$T_a = \sum_{i=1}^n t_i \quad (5)$$

An analysis of the necessary positioning of the baseplate results in a tolerance of 1.95 mm . A stepper motor is chosen, capable of applying a maximum feed force F_{feed} of 242 N . The tolerance T_a according to Equation 5 is 0.138 mm and is well below the necessary positioning tolerance.

C. Gasket Design

According to Ashby [3] and Trutnovsky [4], possible seal materials are plastics and ceramics. Seals made of elastomers, have good sealing properties, but poorer abrasion resistance. In addition, the seal

works in a mass spectrometer, which severely limits the choice of plastics due to degassing. The gasket is made of PTFE and is produced on a lathe. To achieve a good sealing effect, surface pressure compensates for unevenness. Since it is a seal with a moving counter surface, the abrasion on the seal increases with higher contact pressure. The aim is to apply only enough pressure to ensure adequate sealing effect. Literature recommends for PTFE-Gaskets a surface pressures, for static gaskets, are between 3.4– 6.9 N/mm²[4] and 15– 90 N/mm² [5].

D. Sealing Tests

For the pressure monitoring, two pressure measurement cells (Pfeiffer Vakuum PKR 360 [6]) are used, see Figure 3. The pressure cells are calibrated by the manufacturer. The first cell is located in the lens system behind the base plate. The second one is in the TOF analyzer.

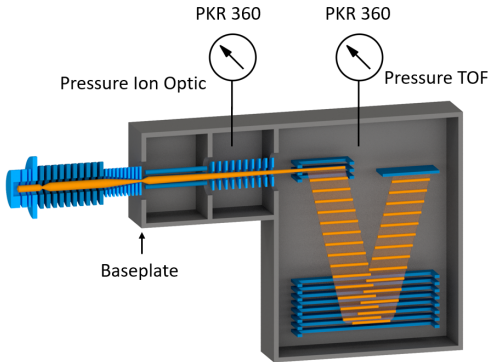


Fig. 3. Positions of the pressure gauges (Symbols: [7]).

The manufactured gasket is evaluated concerning its sealing ability. Following tests are carried out:

- Contact Force Test
- Durability Test
- Pressure Differential Test

The contact force test and the durability test, are done with the baseplate and no reaction chambers,

which are an additional leakage source. At the bottom of the baseplate, two cones are manufactured for a smooth transition edge, as it slides on the gasket, see Figure 4. If the cone slides from atmosphere into the vacuum, a dead volume is introduced, causing a pressure peak. In the contact force test, the spring

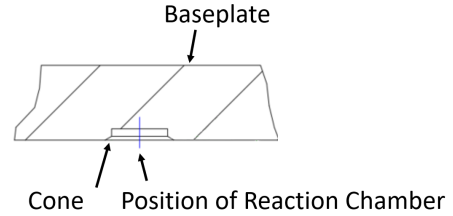


Fig. 4. Cone in the baseplate (section cut) without the reaction chamber. Position on the bottom side, see Figure 2.

deflection is varied and the pressure is measured. The base plate is moved on the gasket. The pressure is measured and the pressure peaks extracted. Five different clamping forces are tested.

For the durability test, the plate is pushed back and forth 100 times, in order to validate tightness and abrasion. The contact pressure through the springs is set to the optimal value (622 N). The pressure peaks from each cycle are extracted using the Signal Processing Toolbox from Matlab. The pressure peaks are of interest, because the vacuum pumps can be damaged.

In the vacuum differential pressure test, the leakage rate of a system can be calculated. The test is carried out in two configurations of the sliding mechanism: When the reaction chamber is in the measurement position, and in the middle/blind position. The influence of the leakage caused by the reaction chamber and the gasket is investigated. The pressure is measured over a defined period of time (30 min) with disconnected pumps. From the pressure increase Δp , the volume V of the vacuum system and the time Δt , the leakage rate Q_L is calculated as follows [8]:

$$Q_L = \frac{\Delta p V}{\Delta t} \quad (6)$$

E. Position Calibration

For the calibration of the drive system, the motor moves the base plate above the subsequent hexapole at a constant speed, see Figure 5. When the reaction

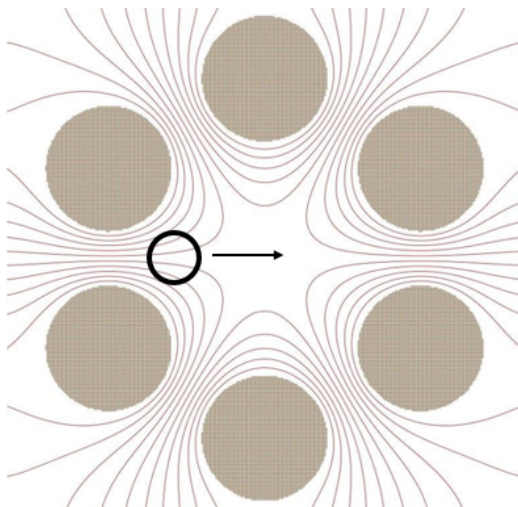


Fig. 5. Position calibration procedure as a top view. The black circle indicates the hole in the baseplate and the direction of movement. Through this hole, the ions move into the hexapole (into the plane of the schematic). [9]

chamber moves over the hexapole, the mass spectrometry signal should increase, reach a maximum and decrease again. Three different masses, of an calibrated gas standard are used. Secondly, the maximum signal of each mass is identified and divided in half $h/2$. Then, the two measuring points closest to half of the maximum value are defined. These two measured values can be assigned to the rising and falling slopes of the signal. The ideal point of the respective mass is then located exactly between these two positions, which equals the half of the width ($w/2$). This process is repeated with each mass and the average of the ideal positions is calculated. The position of the signal maximum cannot be used, as signal noise can falsify this position. The search procedure of the ideal position is shown in Figure 6

and is done in Matlab:

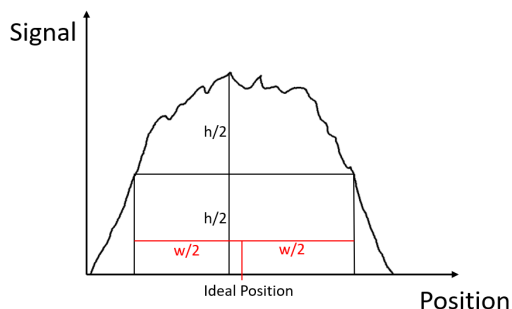


Fig. 6. Position finding procedure as schematic.

F. Exchange Process of the Reaction Chambers

For later operation, the instrument is extended with valves in order to execute the exchange process. The circuit diagram is shown in Figure 7. Two exchange

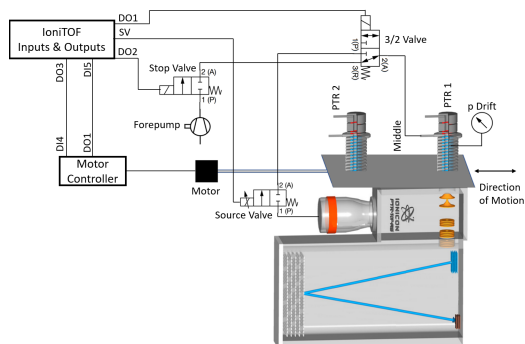


Fig. 7. Schematic circuit plan used for the automation for the exchange of "PTR 1" to "Middle" and back (Symbols: [10], [11]).

processes are realised: fully automatic exchange of one reaction chamber (out- and in-coupling from vacuum to atmosphere) and exchange between two chambers. For the exchange between two chambers, hoses and cables are change manually.

During the automatic exchange of one chamber, the repeatability of the positioning of the motor is measured. Additionally, a mass spectrum before and after a exchange cycle is measured. The repeated signal intensity of three calibration masses is compared. During the process, the cycle time is measured and the pressure is monitored. This should not exceed 1×10^{-4} mbar in the lens system. For the exchange between two chambers, the time is measured.

III. RESULTS

A. Pressure with Different Clamping Forces

Figure 8 shows the maximum measured pressures with different contact forces. As expected, the grey line (cone leaving the vacuum) is lower than when the cone enters the vacuum with the dead volume. There is a clear tendency of the gasket to seal better with increasing contact pressure. While between 454 N and 566 N, the pressure peak reduces by 46 % to 50 %, the change from 566 N to 677 N is only between 33 % and 35 %. The increase of the contact pressure from 789 N to 901 N shows no significant pressure reduction.

As the sealing effect does not improve significantly at a contact pressure higher than 677 N, further operation is done with a contact pressure of 622 N. Therefore, abrasion should be reduced by applying a minimum possible contact pressure.

B. Pressure Differential Test

Figure 9 shows the measured differential pressure with and without the reaction chamber.

The black line is the configuration with the reaction chamber and is clearly steeper than the grey line, i.e. without the reaction chamber in the middle/blind position. This effect is to be expected, as the reaction chamber consists of several components that are sealed to each other, resulting in more leakage possibilities. Thus, the grey line, the differential pressure in the middle position, can be assigned to the leakage of the entire TOF and the constructed PTFE seal. The increased leakage, which can be seen in

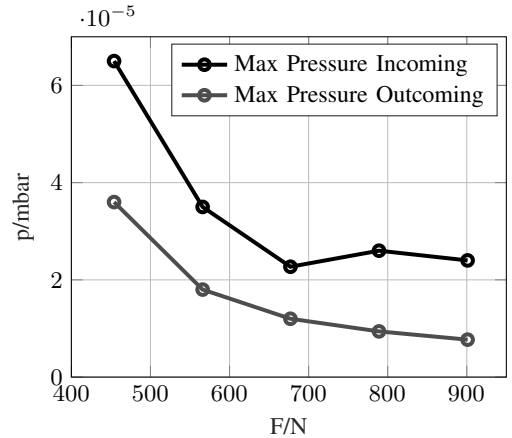


Fig. 8. Maximum measured pressure compared to the clamping force. The black line shows the maximum pressure when the cone enters the vacuum and the gray line when it leaves the vacuum.

the comparison of the black and the grey line, can therefore only be attributed to the reaction chamber, as the rest of the system is unchanged.

For both configurations, the leakage rate, according to Equation 6, results in the following values:

- Q_L with reaction chamber: 0.74 mbar L/s
- Q_L without reaction chamber: 0.087 mbar L/s

Without the reaction chamber, the leakage rate is 8.5 times smaller than with the reaction chamber. It is clear, that the reaction chamber is a weak point in the leak tightness of the entire mass spectrometer. On the other side, the designed gasket shows sufficient sealing performance.

C. Durability Test

Figure 10 shows the maximum pressure peaks, in the lens system. During the whole test, the clamping force is constant. A close look at the measured values shows, that the range of the pressure peaks becomes smaller after cycle 70. Resulting in a more constant sealing effect. While the pressure peaks vary in a range of 4.57×10^{-5} mbar up to cycle 70, the range

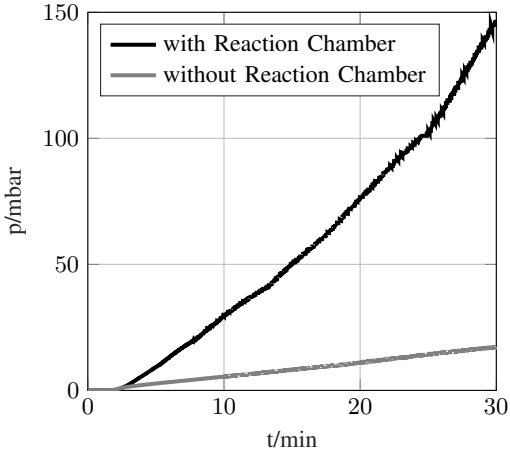


Fig. 9. Pressure difference with and without the reaction chamber.

between cycles 70 and 100 is only 2.98×10^{-5} mbar. After disassembly of the gasket and visual inspection, no abrasion can be seen on the gasket and in the instrument.

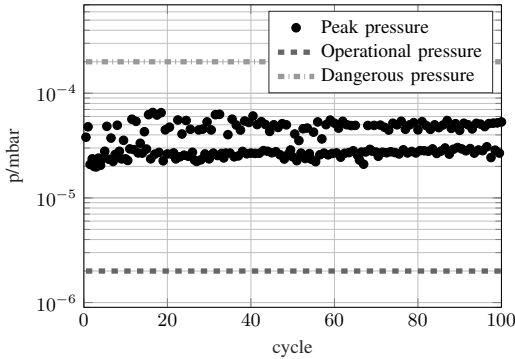


Fig. 10. Peak pressure inside the lens system during 100 push/pull cycles.

D. Position Calibration

Figure 11 shows the signal intensities of three different masses, over a range of 11 mm and constant feed rate.

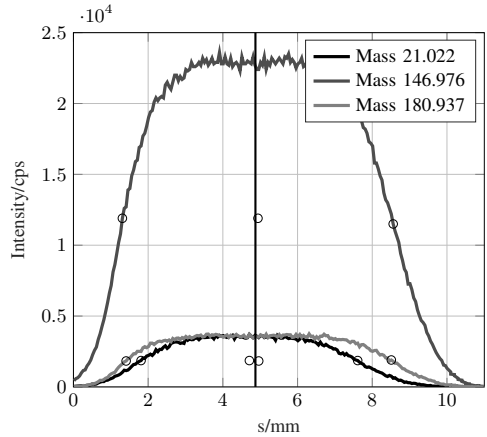


Fig. 11. Signal intensities in cps (counts-per-second), during constant feed of the reaction chamber. The masses have different concentrations in the calibration gas. The vertical line is the evaluated optimal position.

The recorded data of the different masses look like a symmetrical bell curve with a plateau. The plateau is caused by the effective potential inside the hexapole. Mass 21.022 is the narrowest curve, and has the smallest plateau, in contrast to the much heavier masses of the calibration gas. As the ionised molecules enter the alternating electric field of the hexapole, light molecules react much more sensitive to position deviations than heavier molecules. Light molecules are more easily lost, due to concentricity deviations, than heavy ones.

Since the curves do not have a clear maximum, but a plateau, a peak fit is applied using the flanks at half height, according to the procedure in Figure 6. The marked circles show the flanks at half height and the the middle of the flanks. The average of the three peak positions, is indicated by the vertical line. With three repetitions, the average of the calculated posi-

tions show a range of 0.0083 mm. Compared to the theoretical tolerance of the system of 0.138 mm, the tolerance does not occur in significant manner. This might be caused by the fact, that all measurements are done in the same feed direction. The axial play of the linear actuator, is not visible.

E. Exchange Process

During operation, the baseplate is positioned five times within a range of 0.10 mm. For later operation, it is of interest, that after an exchange process, the produced signal intensity, measured at the mass spectrum, is reproducible (within a target range of $\pm 5\%$). With three calibration masses, this target is just reached at three out of ten repetitions. In some cycles one mass, in other cycles even all masses have a higher deviation. At two cycles, the deviation is around 12% less than before the executed cycle. Additionally, the ion source of the reaction chamber did not switch on immediately.

The duration of one cycle, for out- and in-coupling of one reaction chamber, fully automated, is 210 s. Included is the time for the stop and the start of the mass spectrometry measurement. Switching between two chambers, changing hoses and cables manually, takes between 11 and 12 min.

During five exchange cycles, the maximum pressure peaks p_{max} are between 5.88×10^{-5} mbar and 6.26×10^{-5} mbar. These values are solidly below the critical value of 1×10^{-4} mbar.

IV. CONCLUSION & PERSPECTIVE

The results show, that the concept works with a PTFE gasket. During tests, the TOF pressure does not rise above 5×10^{-6} mbar and the pressure in the lens system, during an exchange process does not exceed above 1×10^{-4} mbar. Although the used seal shows a significantly higher asymptotically achievable final pressure, compared to the standard configuration, it is still in a completely acceptable range.

The test with different contact forces, allows to apply the minimum possible contact pressure, that

still guarantees enough tightness. The durability test shows, that tightness is maintained over 100 cycles. Abrasion and poorer sealing ability is not observed. It can be concluded that the service life of the seal extends far beyond the 100 cycles.

With regard to tightness and abrasion, a polished surface essentially improves the performance of the seal. Experiences, regarding the test setup, also show, that the flatness of the sealing surface and of the base plate have an influence on the sealing ability. Important future tests will be using a machine grinded an polished stainless steel baseplate.

The contact pressure concept, using guidance rails and compression springs is durable and functional. During the durability test, the determined contact force remained the same. No readjustment was necessary. Nevertheless, it would be possible, to achieve increased tightness by increasing the contact pressure in the static measuring mode and only using a lower contact pressure when the base plate moves. A method with an adjustable clamping force can be developed in the future.

The tolerance analysis of the drive system shows, that even in the worst summation, reliable positioning is possible. Measurements of the repeatable position of the baseplate almost correspond to the theoretical tolerance of the system. The used drive system is beyond the requirements and offers a low risk of falsification of the actual measurement.

Additionally, the calibration of the drive system by the signal during constant forward movement works. The repetitions show small deviations of the calculated position (0.0083 mm range), which proves the applicability of the procedure.

As a result of the project, changing between two chambers is already possible in approximately 12 min, with hoses and cables being changed manually. The coupling in and out of a single reaction chamber is already possible fully automated and can be done in 3 min and 30 s and shows the potential of a automatic exchange.

During some tests, the ion source of the reaction chamber did not switch on immediately, as well as

impurities occurred in the mass spectrum. Also, the change of the signal intensity, lower than $\pm 5\%$, after a exchange cycle, is not always possible. Here, further investigations are necessary.

Nevertheless, the overall performance is highly satisfying. Counter tests with the standard baseplate show, that measurement results are almost equal with the standard configuration and the developed vacuum-tight exchange mechanism.

ACKNOWLEDGMENT

The author would like to thank Stefan Feil, Paul Mutschlechner, Jakob Heller, Sebastian Repetzki, Lukas Märk, the company Ionicon Analytik GmbH and Jasmin Höglauer for their support during the project.

REFERENCES

- [1] W. Lindinger, A. Hansel, and A. Jordan, "On-line monitoring of volatile organic compounds at pptv levels by means of proton-transfer-reaction mass spectrometry (ptr-ms) medical applications, food control and environmental research," *International journal of mass spectrometry and ion processes*, vol. 173, no. 3, pp. 191–241, 1998.
- [2] S. von Praun, "Toleranzanalyse nachgiebiger Baugruppen im Produktentstehungsprozess," *Doktorarbeit, Universität München*, 2002.
- [3] M. F. Ashby, *Engineering materials 1: an introduction to properties, applications and design*, fifth edition. ed. Butterworth-Heinemann, 2018.
- [4] K. Trutnovsky, *Berührungsdichtungen an ruhenden und bewegten Maschinenteilen*, 2nd ed. Springer-Verlag Berlin Heidelberg GmbH, 1975.
- [5] "Produktdokumentation Flachdichtungen," Available at <https://www.klinger-kempchen.de/wp-content/uploads/2020/07/Flachdichtungen.pdf>, [Online; Date 11.07.2022].
- [6] "PKR 360, low current, DN 25 ISO-KF," Available at <https://www.pfeiffer-vacuum.com/en/products/measurement-analysis/measurement/activeline/activeline-gauges/15899/pkr-360-low-current-dn-25-iso-kf>, [Online; Date 06.08.2022].
- [7] "Datei:Symbol Manometer," Available at https://de.m.wikipedia.org/wiki/Datei:Symbol_Manometer.svg, [Online; Date 06.08.2022].
- [8] K. Jousten, *Handbuch Vakuumtechnik*, 12th ed. Springer Reference Technik, 2018.
- [9] S. Feil, "A Novel Atmospheric Pressure Interface Time-of-Flight (APi-TOF) Instrument for Simultaneous Detection of Positive and Negative Ions," Talk, 2018, European Geosciences Union, Vienna.
- [10] A. Jordan, S. Haidacher, G. Hanel, E. Hartungen, L. Märk, H. Seehauser, R. Schottkowsky, P. Sulzer, and T. Märk, "A high resolution and high sensitivity proton-transfer-reaction time-of-flight mass spectrometer (ptr-tof-ms)," *International Journal of Mass Spectrometry*, vol. 286, no. 2, pp. 122–128, 2009.
- [11] "Bürkert Fluid Control Systems," Available at <https://www.buerkert.at/de>, [Online; Date 31.07.2022].



Johannes Höglauer is student at MCI Innsbruck/Austria. Additionally, he works at Ionicon Analytik GmbH in the R&D department for the mechanical development of mass spectrometers.

Ausbildungsplattform Robotik - Regelung von Industrierobotern

Dominik Huber, Benjamin Massow (Betreuer) und Andreas Mehrle (Betreuer)

Kurzfassung—Robotik ist eine wissenschaftliche Disziplin, in der alle Teilbereiche der Mechatronik bewältigt werden müssen. In dieser Arbeit wird die Umsetzung einer Ausbildungsplattform für Roboter mit Parallelkinematik für den Einsatz auf Hochschulen beschrieben. Lehrinhalte wie Modellbildung und Reglerentwurf sollen schrittweise anhand eines linearen Delta Roboters vermittelt werden. Der Schwerpunkt liegt dabei auf der Berechnung der Dynamik mittels Lagrange-Funktion und Implementierung von Regelungsalgorithmen wie Motion Control. Für die Modellierung des Antriebstranges wird ein grey box model verwendet, um Einflüsse wie Reibung wahrheitsgetreu im Modell abzubilden. Sämtliche Theorien und Annahmen werden durch die Umsetzung an dem virtuellem und realen Robotermodell validiert. Der lineare Delta Roboter wird dabei durch eine speicherprogrammierbare Steuerung mit Feldbusklemmen angesteuert. Die Regelung wird mittels automatischer Codegenerierung vom virtuellem auf das reale Robotermodell übertragen. Nach der mathematischen Beschreibung des Roboters werden modellbasierte Regelstrategien in dieser Arbeit umgesetzt. Die Lehrinhalte sind so aufbereitet, dass angehende Ingenieure einer Robotik Lehrveranstaltung für parallele Kinematiken auch ohne externe Hilfe folgen können.

Schlagwörter—Linear Delta Roboter, Ausbildung, Parallele Kinematik, Motion Control

I. EINLEITUNG

DIESER Artikel beschreibt die Umsetzung einer Ausbildungsplattform zur Lehre von Regelungstechnik in Bezug auf parallele Robotik. Dabei wird

D. Huber studiert am Studiengang Mechatronik, MCI, Innsbruck, Österreich, e-mail: dom.huber@mci4me.at.

B. Massow und A. Mehrle lehren am Studiengang Mechatronik, MCI, Innsbruck, Österreich.

Manuskript eingereicht am 11. August 2022; revidiert am 12. August 2022.

der Roboter DELTA ROBOT [1] des Herstellers IGUS mit Parallelkinematik verwendet wie in Abb. 1 gezeigt.

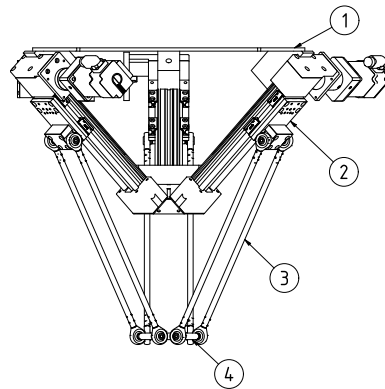


Abbildung 1. Grundstellung des Delta Roboters mit fixierter Plattform (1), Schlitten auf Linearführung (2), Arm (3) und mobile Plattform (4).

Es sollen Themen behandelt werden aus denen die angehenden Ingenieure auch noch in 20 Jahren einen Nutzen ziehen können, wie in [2] beschrieben. Unter anderem sollen Sensoren und Aktoren, lineare und nichtlineare Regelungstheorie gelehrt werden. In [3] wird eine Robotik Toolbox MATLAB entwickelt. Diese ist jedoch nur für serielle Kinematiken geeignet. Die Toolbox von P. Corke sollte außerdem erwähnt werden, welche in [4] gut dokumentiert ist jedoch auch keine parallelen Kinematiken zulässt. Das Feedback von Studenten wird in [5] bei der Verwendung von interaktiver Software, virtuellen

Laboren und E-Learning verwendet. Dabei wurde die Verwendung von Simulationen, Projektarbeiten und Förderung von internationalen Turnieren positiv bewertet.

Es gibt auch einige professionelle/kommerzielle Angebote zur Robotersimulation wie ROBOTSTUDIO [6], TECNOMATIX [7] und EASY-ROB [8]. Im Allgemeinen fehlt es bei diesen Lösungen an der nötigen Anpassungsfähigkeit, um diese sinnvoll als Ausbildungsplattform einzusetzen.

Der Student soll in dieser Ausbildungsplattform folgende Arbeitsschritte durchführen:

- Berechnung der Kinematik
- Berechnung der Dynamik
- Validierung der Kinematik und Dynamik
- Reibungsmodellierung
- Berechnung der Trajektorie
- Regelung simulieren in SIMULINK/SIMSCAPE
- Regelung implementieren in SPS

Diese Arbeitsschritte können dabei in Gruppen in Form eines Projektes ausgearbeitet werden. Dieser Artikel knüpft dabei an die Masterarbeit [9] an. Der darin erarbeitete mechanische Aufbau und elektrischen Komponenten werden in dieser Arbeit verwendet. Auch werden theoretische Grundlagen der Robotik wie direkte und indirekte Kinematik in [9] näher beschrieben und daher in dieser Arbeit nicht behandelt.

II. METHODEN

A. Beschreibung des Roboters

Die Mobilität/Freiheitsgrade des Roboters μ kann laut [10] mittels der Kutzbach-Grübler Gleichung

$$\mu = \lambda n_m - (5 j_1 + 4 j_2 + 3 j_3) \quad (1)$$

berechnet werden. Im dreidimensionalen Raum ist die Anzahl der Freiheitsgrade sechs, deshalb entspricht $\lambda = 6$. Mittels der Anzahl der beweglichen Glieder n_m und der Anzahl der Gelenke j_k kann die Mobilität berechnet werden. Index k entspricht dabei der relativen Mobilität des Gelenkes, zum Beispiel

$k = 3$ für ein sphärisches Gelenk. Daraus resultiert für den in Abb. 2 dargestellten Manipulator

$$\mu = 6 \cdot 10 - (5 \cdot 3 + 3 \cdot 12) = 9. \quad (2)$$

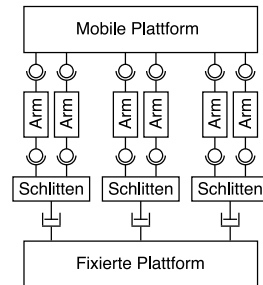


Abbildung 2. Schematische Darstellung der Kinematik mit zwölf sphärischen und drei prismatischen Gelenken.

Die sechs Arme haben einen rotativen Freiheitsgrad um deren Stabachse, der sich nicht beobachten oder steuern lässt. Die restliche drei Freiheitsgrade aus Gl. (2) beziehen sich auf die mobile Plattform, welche sich durch die Bewegung der drei prismatischen Gelenke steuern lässt. Die mobile Plattform ist dabei stets parallel zur fixierten Plattform und es kommt zu keiner Rotation zwischen beiden Plattformen wie in [11] beschrieben.

Um eine Simulation des Robotermodells in SIMULINK/SIMSCAPE stabil durchführen zu können, muss die freie Rotation der Arme unterbunden werden. Dies wird erreicht indem die sechs sphärischen Gelenke, die Schlitten und Arm verbinden, durch Universalgelenke ersetzt werden. Die Mobilität sinkt dabei auf

$$\mu = 6 \cdot 10 - (5 \cdot 3 + 4 \cdot 6 + 3 \cdot 6) = 3. \quad (3)$$

Das Trägheitsmoment $J_{arm,z}$ des Arm um deren lokale z -Achse wird aufgrund des kleinen Einflusses in der Berechnung der Dynamik vernachlässigt. Lokales Koordinatensystem ist definiert wie in Abb. 3 dargestellt.

Der Roboter wird in Abb. 4 schematisch dargestellt. Jeweils zwei Arme die im realen Roboter-

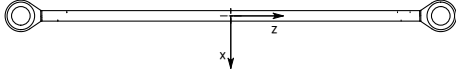


Abbildung 3. Arm des Roboters mit lokalem Koordinatensystem mit Ursprung im Schwerpunkt des Objektes.

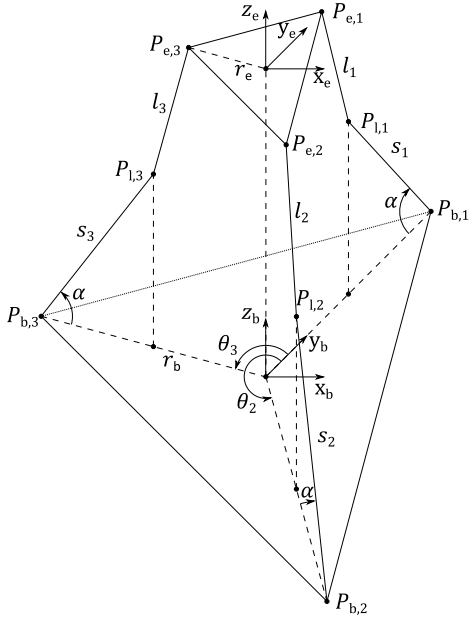


Abbildung 4. Geometrie des Roboters mit dem Basiskoordinatensystem mit Index b und Koordinatensystem der mobilen Plattform mit Index e .

modell mit einem Schlitten verbunden sind, werden durch einen virtuellen Arm l_k für $k = 1, 2, 3$ ersetzt. Arm l_k befindet sich dabei zwischen den beiden realen Armen. Punkte $P_{b,k}$ sind die Drehpunkte der virtuellen Arme l_k mit dem Schlitten, sobald alle Schlitten in ihrer Grundstellung sind. Die Grundstellung des Roboters ist in Abb. 1 dargestellt. Dabei sind die Schlitten 3 mm von deren mechanischen Anschlag entfernt. Im Schwerpunkt dieses von den Punkten $P_{b,k}$ aufgezogenen Dreiecks bildet den Ursprung des Basiskoordinatensystems. Selbes gilt für

das Koordinatensystem der mobilen Plattform, bei der Punkte $P_{e,k}$ die Drehpunkte der virtuellen Arme mit der mobilen Plattform darstellen. Die linearen Achsen des Roboters, durch welche der Roboter gesteuert wird, werden als s_k bezeichnet.

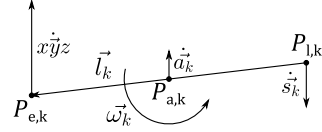


Abbildung 5. Definition translatorische und rotatorische Geschwindigkeiten des Roboterarm

Abb. 5 zeigt schematisch einen Ausschnitt aus Abb. 4 mit einem Arm der Länge l_k . Punkt $P_{a,k}$ stellt den Schwerpunkt des Arms dar und befindet sich mittig zwischen den beiden Punkten $P_{e,k}$ und $P_{l,k}$. Die Geschwindigkeit des Armschwerpunktes \dot{a}_k wird durch

$$\dot{a}_k = \frac{x\dot{y}z + \dot{s}_k}{2} \quad (4)$$

berechnet. Die Winkelgeschwindigkeit $\vec{\omega}_k$ beschreibt die Drehung des Arms um dessen Schwerpunkt. Durch

$$\vec{\omega}_k = \frac{4}{\|\vec{l}_k\|^2} \left(\vec{l}_k \times (x\dot{y}z - \dot{a}_k) \right) = \frac{4}{\|\vec{l}_k\|^2} \left(\vec{l}_k \times (\dot{a}_k - \dot{s}_k) \right) \quad (5)$$

wird diese Winkelgeschwindigkeit aus zweier Geschwindigkeiten berechnet. Dies wird mit den bekannten Koordinatensystemen $x\dot{y}z$ und \vec{s} ausgedrückt durch

$$\vec{\omega}_k = \frac{1}{\|\vec{l}_k\|^2} \left(\vec{l}_k \times (x\dot{y}z - \dot{s}_k) \right) \quad (6)$$

$$\|\vec{\omega}_k\| = \frac{\|x\dot{y}z - \dot{s}_k\|}{\|\vec{l}_k\|}$$

B. Berechnung der Dynamik

Der Manipulator wird zum Berechnen der Dynamik simplifiziert als System starrer Körper angenommen. Zur Berechnung des dynamischen Modells wird

die Lagrange-Theorie verwendet. Die Lagrange'sche Mechanik hat gegenüber der Newton'schen Mechanik Vorteile bei der Berechnung von Bewegungsgleichungen bei Systemen mit vielen Zwangsbedingungen. Die Lagrange-Theorie ist laut [12] wegen ihrer einfachen Struktur und ihrer direkten Verbindung zur Newton'schen Theorie der Hamilton'schen Mechanik vorzuziehen. Die Lagrange-Funktion für Systeme mit holonomen Zwangsbedingungen lässt sich durch

$$L = T - U \quad (7)$$

ausdrücken mit T für kinetische und U für potenzielle Energie des betrachteten Systems. Die kinetischen Energien sind in diesem Fall rotatorisch und translatorischer Art. Die potenzielle Energie ist die Lageenergie der Körper im Gravitationsfeld der Erde. Die Energien werden dabei in den generalisierten Koordinaten

$$\vec{s} = \begin{bmatrix} s_1 \\ s_2 \\ s_3 \end{bmatrix} \quad (8)$$

und

$$x\vec{y}z = \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} x_b - x_e \\ y_b - y_e \\ z_b - z_e \end{bmatrix} \quad (9)$$

ausgedrückt.

Die Lagrange-Gleichung wird in 2 Teile aufgeteilt. In die Gleichung mit den Schlittenvariablen

$$\vec{\tau}_s = \frac{d}{dt} \left(\frac{\partial L}{\partial \dot{\vec{s}}} \right) - \frac{\partial L}{\partial \vec{s}} \quad (10)$$

bei welchen die berechneten Kräfte $\vec{\tau}_s$ bereits an den Schlitten angreifen und in

$$\tau_{xyz} = \frac{d}{dt} \left(\frac{\partial L}{\partial x\dot{y}z} \right) - \frac{\partial L}{\partial x\vec{y}z} \quad (11)$$

bei denen die Kräfte τ_{xyz} am Endeffektor angreifen.

Die beiden Kraftvektoren $\vec{\tau}_s$ und τ_{xyz} können laut [13], dann zu der gesamt benötigten Kraft an den Schlitten $\vec{\tau}$ summiert werden mit

$$\vec{\tau} = \vec{\tau}_s + \mathbf{J}^T \tau_{xyz}. \quad (12)$$

\mathbf{J} stellt dabei die Jacobimatrix dar.

Dabei erhält man laut [13] für einen starren Roboter ein dynamisches Modell der Form

$$\vec{\tau} = \mathbf{M}(\vec{s}) \ddot{\vec{s}} + \vec{c}(\vec{s}, \dot{\vec{s}}). \quad (13)$$

$\mathbf{M}(\vec{s})$ stellt dabei die Trägheitsmatrix und $\vec{c}(\vec{s}, \dot{\vec{s}})$ Coriolis-, Zentrifugal- und Gravitationseffekte dar.

C. Validierung der Kinematik und Dynamik

Die kinematischen und dynamischen Berechnungen des Roboters werden validiert durch die Methode, welche in Abb. 6 abgebildet ist. Das Modell

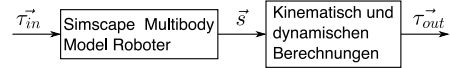


Abbildung 6. Schematische Übersicht zur Methode der Validierung der kinematischen und dynamischen Berechnungen.

des Roboters wird als Simscape Multibody importiert. Hierbei werden die Bewegungen der einzelnen Glieder durch deren Trägheiten und Kräfte mittels Newton'scher Gesetze berechnet.

Der Kraftvektor $\vec{\tau}_{in}$ ist dabei eine Kraft die direkt in Bewegungsrichtung der Schlitten angreift. Die gemessene Variable \vec{s} aus diesem Modell wird als Eingang für die kinematische und dynamische Berechnung verwendet. Mittels dieser Berechnung wird wiederum der Kraftvektor $\vec{\tau}_{out}$ berechnet, der für diese Bewegungen notwendig ist. Ein Abgleich zwischen $\vec{\tau}_{in}$ und $\vec{\tau}_{out}$ validiert die kinematischen und dynamischen Berechnungen.

D. Reibungsmodellierung

Reibungseffekte sind kritisch für Roboter, da sie beispielsweise zu Positionierungsfehlern führen. Die Reibung muss daher kompensiert werden, um diesen Einfluss zu eliminieren. Die Reibkraft, welche auf den Schlitten in Bewegungsrichtung wirkt, wird angenommen als

$$\tau_{f,k} = \tau_{f,Sch,k} + \tau_{f,Arm,k}. \quad (14)$$

$\tau_{f,Sch}$ beschreibt den Widerstand des Schlittens auf Linearführung und $\tau_{f,Arm}$ beschreibt den Widerstand durch die Armgelenke. Die Modellbildung der Reibungen dieser beiden Teile wird durch Messungen am realen Roboter durchgeführt.

Der Widerstand des Schlittens gegen eine Bewegung wird durch

$$\tau_{f,Sch,k} = \tau_{f,Sch,pos,k}(s_k, \text{sign}(\dot{s}_k)) + \tau_{f,Sch,vel,k}(\dot{s}_k) \quad (15)$$

modelliert. Dabei ist $\tau_{f,Sch,pos,k}$ eine positionsabhängige Kraft, die zudem von der Bewegungsrichtung abhängt. Diese Annahme wird aufgrund des eingesetzten Zahnriemens getroffen. $\tau_{f,Sch,vel,k}$ ist rein abhängig von der Schlittengeschwindigkeit \dot{s}_k . Durch den Aufbau werden auch Reibungen aller Komponenten entlang des Antriebstranges berücksichtigt. Polygoneffekte werden bei diesen Messungen nicht berücksichtigt, es wird ein konstantes Übersetzungsverhältnis angenommen.

Zur Bestimmung der Kraft $\tau_{f,Sch,pos,k}$ wird eine konstante Schlittengeschwindigkeit von 1 mm s^{-1} verwendet. Die Messwerte werden anschließend in der Regelung in Form einer LUT mit einer Positionsauflösung von $0,1 \text{ mm}$ verwendet.

Die Messwerte zur Bestimmung der Kraft $\tau_{f,Sch,vel,k}$ wird durch zwei unterschiedliche Methoden bestimmt:

- $\dot{s}_k \leq 20 \text{ mm s}^{-1}$: Konstante Geschwindigkeit \dot{s}_k , Kraft am Schlitten τ_k arithmetisch gemittelt
- $\dot{s}_k > 20 \text{ mm s}^{-1}$: Konstante Kraft am Schlitten τ_k , Geschwindigkeit \dot{s}_k arithmetisch gemittelt

Die Kraft $\tau_{f,Sch,vel,k}(\dot{s}_k)$ wird anschließend durch ein Polynom dritter Ordnung gefittet. Die verwendete Gleichung

$$\tau_{f,Sch,vel,k} = a_{0,k} \text{sign}(\dot{s}_k) + a_{1,k} \dot{s}_k + a_{2,k} \text{sign}(\dot{s}_k) \dot{s}_k^2 + a_{3,k} \dot{s}_k^3 \quad (16)$$

wird dabei in [14] beschrieben.

Zur Bestimmung des Widerstandes den die Armgelenke der Bewegung entgegensetzen wird ein konstantes Reibmoment $\|M_r\|$ an den Gelenken $P_{e,k}$ und $P_{l,k}$ angenommen. Dieses Moment wirkt der

Drehung ω_k des Arms entgegen. Durch Ansetzen des

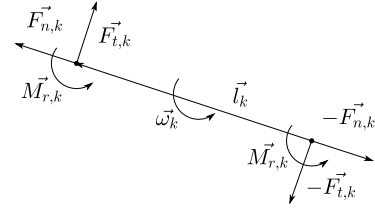


Abbildung 7. Definition der angreifenden Momente und Kräfte am Roboterarm.

Momentengleichgewichts am Arm in Abb. 7 folgt

$$F_{t,k} = \frac{1}{\|\vec{l}_k\|^2} (\vec{l}_k \times 2\vec{M}_{r,k}). \quad (17)$$

Die Kraft $F_{n,k}$ wird anschließend durch Ansetzen des Kräftegleichgewichts am Endeffektor und der Berechnung des linearen Gleichungssystems gelöst.

Mit den beiden Kräfte $F_{n,k}$ und $F_{t,k}$ wird anschließend die Widerstandskraft in Bewegungsrichtung des Schlittens $\tau_{f,Arm,k}$ berechnet.

E. Regelung

Durch die Berechnung der Roboter Kinematik und Dynamik kann dieser durch modellbasierte Regelungsstrategien angesteuert werden. Dies ist besonders bei Bewegungen nötig die schnell und präzise durchgeführt werden müssen. Mit der Ausbildungsplattform können verschiedenste Regelungsstrategien wie Motion Control umgesetzt werden.

Dabei werden diese zuerst am virtuellem Modell des Roboters in SIMULINK/SIMSCAPE evaluiert. Anschließend wird der Regler mittels SIMULINK PLC CODER in strukturierten Text compiliert und in TWINCAT importiert. Hardwarelimits des Manipulators werden durch den Treiber auf der SPS überprüft. Bei Überschreitung dieser Limits durch den Regler wird ein Fehler gemeldet und die Anlage gestoppt.

Die Aufgabe von Motion Control ist die Umsetzung einer Referenztrajektorie $xy\dot{z}_r$ am Manipulator. Es wird Motion Control durch Feedback Linearisierung [15], [16] eingesetzt. Der Vorteil einer Feedback

Linearisierung ist, dass man zur Reglerauslegung sämtliche Methoden für lineare Systeme verwenden kann. Das nichtlineare Modell kann so zur Auslegung der Regelung im simplifizierten Fall als System zweiter Ordnung angenommen. In

$$\begin{aligned}\vec{\tau} &= \mathbf{M}(\vec{s}) (x\ddot{y}z_r + \vec{u}) + \vec{c}(\vec{s}, \dot{\vec{s}}) + \vec{\tau}_f \\ \vec{u} &= \mathbf{K}_p \vec{e} + \mathbf{K}_d \dot{\vec{e}}\end{aligned}\quad (18)$$

wird beispielsweise ein PD-Regler verwendet. Die Regelabweichung zwischen den eigentlichen Endeffektorposition $x\vec{y}z$ und der Referenz $x\vec{y}z_r$ wird berechnet durch

$$\vec{e} = x\vec{y}z_r - x\vec{y}z. \quad (19)$$

Die PD-Regelung muss nur dann eingreifen, wenn es zu einer Diskrepanz zwischen dem realen Roboter und dessen Modell kommt.

III. ERGEBNISSE

A. Dynamik

Durch die Methode beschrieben in Unterabschnitt II-B kann die Dynamik des Roboters berechnet werden. Das Ergebnis ist wird als Matrixmultiplikation mit

$$\vec{a} = \begin{bmatrix} \ddot{\vec{s}} \\ x\ddot{y}z \\ g \end{bmatrix} \quad (20)$$

ausgedrückt mit der Erdbeschleunigung g . Die folgenden Matrizen enthalten die Masse der mobile Plattform m_e , die Masse des Schlittens m_s , die Masse des Arm m_a und das Trägheitsmoment des Arms J_a . J_a ist dabei das Trägheitsmoment um die x-Achse laut Abb. 3. Geometrische Größen können aus Abb. 4 entnommen werden. Der Vektor $\vec{\tau}_s$ wird zur Matrizenmultiplikation

$$\begin{aligned}\vec{\tau}_s &= \begin{bmatrix} \tau_{s,1} & 0 & 0 \\ 0 & \tau_{s,2} & 0 \\ 0 & 0 & \tau_{s,3} \end{bmatrix} \\ &\quad \begin{bmatrix} \tau_{s,1,4} & \tau_{s,1,5} & \tau_{s,1,6} & \tau_{s,1,7} \\ \tau_{s,2,4} & \tau_{s,2,5} & \tau_{s,2,6} & \tau_{s,2,7} \\ \tau_{s,3,4} & \tau_{s,3,5} & \tau_{s,3,6} & \tau_{s,3,7} \end{bmatrix} \cdot \vec{a} \quad (21)\end{aligned}$$

vereinfacht, mit den Einträgen

$$\begin{aligned}\tau_{s,k} &= m_s + \frac{m_a}{4} + \frac{J_a}{l^2} \\ \tau_{s,k,4} &= -\frac{\cos(\alpha) \sin(\theta_k) (4 J_a - l^2 m_a)}{4 l^2} \\ \tau_{s,k,5} &= \frac{\cos(\alpha) \cos(\theta_k) (4 J_a - l^2 m_a)}{4 l^2} \\ \tau_{s,k,6} &= -\frac{\sin(\alpha) (4 J_a - l^2 m_a)}{4 l^2} \\ \tau_{s,k,7} &= -\frac{\sin(\alpha) (m_a + 2 m_s)}{2}.\end{aligned}\quad (22)$$

Der Vektor $\vec{\tau}_{xyz}$ wird zur Matrizenmultiplikation

$$\begin{aligned}\vec{\tau}_{xyz} &= \begin{bmatrix} \tau_{xyz,1,1} & \tau_{xyz,1,2} & \tau_{xyz,1,3} \\ \tau_{xyz,2,1} & \tau_{xyz,2,2} & \tau_{xyz,2,3} \\ \tau_{xyz,3,1} & \tau_{xyz,3,2} & \tau_{xyz,3,3} \\ \tau_{xyz,1} & 0 & 0 & 0 \\ 0 & \tau_{xyz,2} & 0 & 0 \\ 0 & 0 & \tau_{xyz,3} & \tau_{xyz,g} \end{bmatrix} \cdot \vec{a} \quad (23)\end{aligned}$$

vereinfacht, mit den Einträgen

$$\begin{aligned}\tau_{xyz,1,k} &= -\frac{\cos(\alpha) \sin(\theta_k) (4 J_a - l^2 m_a)}{4 l^2} \\ \tau_{xyz,2,k} &= \frac{\cos(\alpha) \cos(\theta_k) (4 J_a - l^2 m_a)}{4 l^2} \\ \tau_{xyz,3,k} &= -\frac{\sin(\alpha) (4 J_a - l^2 m_a)}{4 l^2} \\ \tau_{xyz,k} &= m_e + \frac{3 m_a}{4} + \frac{3 J_a}{l^2} \\ \tau_{xyz,g} &= -\frac{3 m_a}{2} - m_e.\end{aligned}\quad (24)$$

B. Validierung Kinematik und Dynamik

Für die Validierung wird der Kraftvektor für die Schlitten $\vec{\tau}_{in}$ konstant gehalten. Dazu kann beispielsweise als Eingangsvektor

$$\vec{\tau}_{in} = \begin{bmatrix} -7,2 \text{ N} \\ -7,5 \text{ N} \\ -7,8 \text{ N} \end{bmatrix} \quad (25)$$

gewählt werden. Diese Kräfte entsprechen dabei nicht dem Gleichgewichtszustand des Roboters in der initialen Konfiguration.

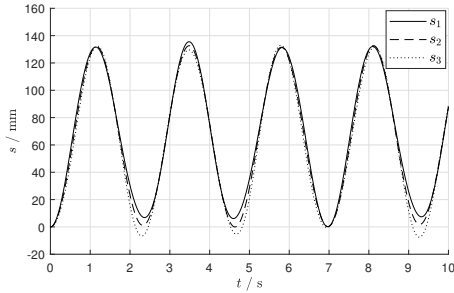


Abbildung 8. Plot der Schlittenpositionen \vec{s} zur Validierung von dynamischer und kinematischer Berechnung.

Beim Betrachten der Schlittenpositionen in Abb. 8 stellt man daher ein Schwingen fest.

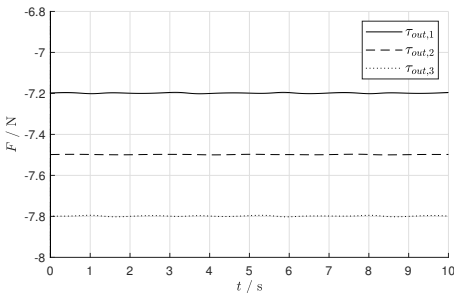


Abbildung 9. Plot des Kraftvektors $\vec{\tau}_{out}$ zur Validierung von dynamischer und kinematischer Berechnung.

Das Ergebnisses der dynamischen Berechnung $\vec{\tau}_{out}$ ist in Abb. 9 im zeitlichen Verlauf dargestellt. Durch den Vergleich der Kraftvektoren III-B $\vec{\tau}_{in}$ und $\vec{\tau}_{out}$ wird daraus geschlossen, dass die dynamischen und kinematischen Berechnungen korrekt durchgeführt wurden.

C. Validierung der Reibungskompensation

Für die Validierung der Reibungskompensation wird die Abweichung von einer Referenztrajektorie als Maßstab verwendet. Als Trajektorie wird ein Polynom siebter Ordnung verwendet. Die Trajektorie wird auf der Hardware während der Laufzeit anhand Stützpunkte berechnet.

Es wird ein Regler mit Motion Control durch Feedback Linearisierung verwendet wie in Gl. (18) beschrieben. Als Regelparameter wird $K_p = 1400 \text{ s}^{-2}$ und $K_v = 200 \text{ s}^{-1}$ verwendet. Dies entspricht einer Eigenkreisfrequenz von $\omega_n = 37,41 \text{ s}^{-1}$ und einem Dämpfungsgrad von $\xi = 2,67$. Die Abtastzeit beträgt $T_s = 1 \text{ ms}$.

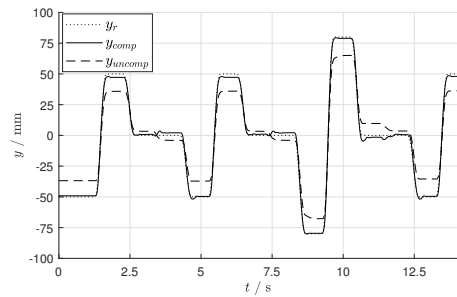


Abbildung 10. Plot der Referenztrajektorie y_r , der gemessenen Trajektorie mit Reibungskompensation y_{comp} und der gemessenen Trajektorie ohne Reibungskompensation y_{uncomp} . y -Koordinaten werden aus der Encoderpositionen berechnet.

Die Messungen sind in Abb. 10 dargestellt. Für den Regler mit der unkompensierten Reibung ergibt sich eine durchschnittliche Abweichung von der Referenztrajektorie von 17,1 mm. Der Regler mit Reibungskompensation hat eine durchschnittliche Abweichung von 2,9 mm. Die Kompensation der Reibung hat einen positiven Einfluss auf die Abweichung von der Referenztrajektorie.

IV. CONCLUSIO

In dieser Arbeit wird eine Ausbildungsplattform erarbeitet, damit zukünftige Ingenieure sich mit Robotik in Bezug auf Parallelkinematik fortbilden können. Auszubildende arbeiten sich dabei durch alle nötigen Schritte, um am Ende eine Regelung als Simulation und am realen Roboter zu erhalten. Die Schritte umfassen dabei die Berechnung der Kinematik und Dynamik, über die Reibungsmodellierung und Trajektorienberechnung bis hin zur Regelungstechnik. Der Programmieraufwand, welcher durch den Ingenieur zu bewältigen ist, ist außerdem klein. Dies wird durch den mitgelieferten Treiber und der automatischen Codegenerierung gewährleistet.

Ein möglicher nächster Schritt ist eine verbesserte Reibungskompensation, beispielsweise mit adaptiver Reibungskompensation mit einem dynamischen Reibungsmodell. Um die Aufgabenstellung zu vereinfachen und die Leistung des Roboters zu verbessern, kann auch die Verwendung von Linearachsen mit Gewindespindel untersucht werden.

LITERATUR

- [1] *Technische Dokumentation, DELTA ROBOT*, igus® GmbH, Nov. 2021.
- [2] R. Bajcsy, "Robotics education today which might last students two or more decades," *Annual reviews in control*, vol. 47, pp. 193–197, 2019.
- [3] A. Gil, O. Reinoso, J. M. Marin, L. Paya, and J. Ruiz, "Development and deployment of a new robotics toolbox for education," *Computer Applications in Engineering Education*, vol. 23, no. 3, p. 443–454, Dec 2014.
- [4] P. I. Corke, *Robotics, vision and control : fundamental algorithms in MATLAB®*. Springer, 2017.
- [5] M. Berenguel, F. Rodríguez, J. C. Moreno, J. L. Guzmán, and R. González, "Tools and methodologies for teaching robotics in computer science & engineering studies," *Computer applications in engineering education*, vol. 24, no. 2, pp. 202–214, 2016.
- [6] ABB Group. Robotstudio. [Online]. Available: <https://new.abb.com/products/robotics/de/robotstudio>
- [7] Siemens Digital Industries Software. Tecnomatix: Siemens software. [Online]. Available: <https://www.plm.automation.siemens.com/global/de/products/tecnomatix/>

- [8] EASY-ROB Software GmbH. Software Modul für Roboter Simulation: Easy-rob. [Online]. Available: <https://easy-rob.com/>
- [9] S. Kollegger, "Entwicklung einer Ausbildungsplattform für parallele Industrieroboterkinematiken," Masterarbeit, MCI Innsbruck, 2021.
- [10] G. Martin, *Getriebelehre eine Theorie des Zwanglaufes und der ebenen Mechanismen*. Berlin Springer, 1921.
- [11] E. A. Baran, O. Ozen, D. Bilgili, and A. Sabanovic, "Unified kinematics of prismatically actuated parallel delta robots," *Robotica*, vol. 37, no. 9, pp. 1513–1532, Feb. 2019. [Online]. Available: <https://doi.org/10.1017/s0263574719000092>
- [12] P. van Dongen, *Klassische Mechanik: Von der Newton'schen Mechanik zur Relativitätstheorie in drei Postulaten*, 1st ed. Berlin, Heidelberg: Springer Berlin Heidelberg, sep 2021.
- [13] S. Briot and W. Khalil, *Dynamics of Parallel Robots*. Springer International Publishing, 2015. [Online]. Available: <https://doi.org/10.1007/978-3-319-19788-3>
- [14] B. Bona, M. Indri, and N. Smaldone, "Nonlinear friction estimation for digital control of direct-drive manipulators," in *2003 European Control Conference (ECC)*. IEEE, 2003, pp. 2685–2690.
- [15] T. R. Kurfess, *Robotics and automation handbook*. CRC Press, 2005.
- [16] J. E. Slotine and W. Li, *Applied Nonlinear Control*. Englewood Cliffs, NJ: Prentice-Hall, 1991.



Dominik Huber studiert am Department für Mechatronik am MCI Innsbruck/Österreich. Im Jahr 2014 erwarb er einen Bachelor-Abschluss mit der Arbeit: „Konzeptionierung und Entwicklung einer Robotik Lehrplattform“.

Entwicklung eines Impulshammers für Schienensysteme

Michael Kofler und Franz-Josef Falkner (Betreuer)

Kurzfassung—Um zu gewährleisten, dass schädliche Frequenzen für die Bausubstanz und störende Frequenzen für die Bewohner, verursacht durch Schienenfahrzeuge, nicht an Nachbargebäude übertragen werden, muss die Übertragungsfunktion zwischen Schienensystemen und Gebäuden ermittelt werden. Somit kann die Lagerung von Schienensystemen und Gebäuden ideal auf die auftretenden Frequenzen abgestimmt werden. Um die Übertragungsfunktion zu ermitteln, wird aktuell, für die Kraftanregung, eine definierte Masse aus einer definierten Höhe fallen gelassen. Das Problem dabei sind die Mehrfachstöße und die dadurch überlagerte Kraftanregung im Frequenzbereich. Ziel der Arbeit ist die Konstruktion und Fertigung eines Impulshammers (Prototyp), welcher die Mehrfachstöße durch eine Bremsvorrichtung verhindert. So kann eine möglichst überlagerungsfreie Kraftanregung im Frequenzbereich erzielt werden. Es gilt auch, den Prototypen zu testen und die Ergebnisse mit der aktuellen Vorgehensweise zu vergleichen. Für den Impulshammer wird eine Vorrichtung verwendet, mit der man die Fallmasse auf die gewünschte Abwurfhöhe hebt. Im Anschluss wird die Masse fallen gelassen und schlägt geführt auf den Untergrund auf. Die beim Aufschlag auftretende Kraft wird mittels Kraftsensor gemessen. Nach dem Aufprall springt die Masse aufgrund des teilelastischen Stoßes wieder vom Untergrund weg. Durch die verbaute Bremsvorrichtung wird der erneute Aufschlag auf den Untergrund verhindert. Die Masse wird mit geringstem Kraftaufwand zum Stillstand gebracht, um einen möglichst kleinen zweiten Impuls in den Untergrund zu induzieren. Weiters wird die Masse des gesamten Impulshammers so weit wie möglich reduziert, um einen einfachen Transport zu gewährleisten. Versuche bestätigen, dass durch das Abbremsen der Masse und Verhindern der Mehrfachstöße die Überlagerung des Kraftsignals im Frequenzbereich auf ein Mini-

mum reduziert wird. Leichte Überlagerungen im Frequenzbereich sind aufgrund der eingebrachten Kraft zufolge des Abbremsens der Fallmasse nach wie vor vorhanden. Jedoch betragen diese Überlagerungen nur noch einen Bruchteil der Ausgangssituation.

Schlagwörter—Mehrfachstöße, Kraftanregung im Frequenzbereich, Schädliche Frequenzen, Schienenfahrzeuge

I. EINLEITUNG

UM zu gewährleisten, dass schädliche Frequenzen für die Bausubstanz bzw. störende Frequenzen für die Bewohnern, verursacht durch Schienenfahrzeuge, nicht an Nachbargebäude übertragen werden, muss die Übertragungsfunktion zwischen Schienensystemen und Gebäuden in einem Frequenzbereich von 0 – 400 Hz (*ÖNORM ISO 14837-1*) ermittelt werden. Dadurch kann die Lagerung von Schienen bzw. Gebäuden ideal auf die auftretenden Frequenzen ausgelegt werden. Aktuell wird eine Masse von 62,5 kg aus einer Höhe von 1,2 m fallen gelassen und zugleich die Beschleunigung der relevanten angrenzenden Gebäude gemessen. Anschließend kann mit der gemessenen Kraft des Impulshammers/ Fallmasse (*INPUT*) und der gemessenen Beschleunigung der Gebäude (*OUTPUT*) die Übertragungsfunktion bestimmt werden. Ein Problem dabei ist, dass die Masse nach dem Rückprall vom Untergrund ein zweites mal auf den Untergrund auftrifft und einen erneuten Impuls (Kraft) indiziert. Durch diese Doppelstöße/ Mehrfachstöße werden gewisse Frequenzbereiche nicht bzw. überlagert angeregt. Somit kann keine präzise Übertragungsfunktion für den gesamt zu

M. Kofler e-mail: km8550@mci4me.at und
F.J. Falkner e-mail: Franz-Josef.Falkner@mci.edu,
MCI, Innsbruck, Österreich.

untersuchenden Frequenzbereich bestimmt werden. Eine weitere Anregungsmethode ist das Einbringen von Schwingungen in den Untergrund mittels eines so genannten *Seismic-Trucks*. Dieser Truck ist mit einem großen *Shaker* ausgestattet und ist in der Lage einen gewissen Frequenzbereich und eine gewisse Kraft anzuregen. Jedoch ist es aufgrund der hohen bewegten Masse nicht möglich, den benötigten Frequenzbereich von 0 – 400 Hz anzuregen. Somit kann auch mit dieser Methode keine verlässliche Aussage über den benötigten Frequenzbereich getätigt werden.

Die Motivation hinter der Arbeit liegt in der Beseitigung der Mehrfachstöße bei gleichzeitiger Erhöhung des Frequenzspektrums. Weiters ist es nicht möglich, den Impuls mit einem händisch geführten Impulshammer einzubringen. Händische Impulshämmer sind in Masse und Kraft begrenzt.

II. METHODEN

Für die Beantwortung der Forschungsfrage wird zuerst die Kraftanregung des Einmassenschwingers betrachtet, denn der Impulshammer regt ebenfalls den Untergrund mit einer Kraft $F(t)$ über Stoßzeit t_s an.

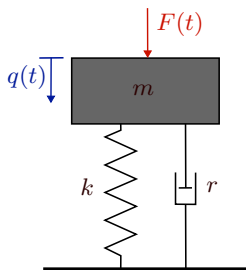


Abb. 1: Kraftanregung des Einmassenschwingers.

Für die harmonische Kraftanregung wird die Kraftanregung $F_0 e^{i\Omega t}$ modelliert und ergibt die Bewegungsgleichung

$$m\ddot{q}(t) + r\dot{q}(t) + kq(t) = F_0 e^{i\Omega t} \quad (1)$$

für den Einmassenschwinger zufolge Kraftanregung (Abb.1). Dabei beschreibt die Variable $m\ddot{q}(t)$ die Trägheitskraft der Masse m und die Variablen $r\dot{q}(t)$ und $kq(t)$ die inneren Kräfte des Dämpfers und der Feder. Die Variable $q(t)$ steht für die *generalisierte Koordinate*, im Fall des Einmassenschwingers für die vertikale Verschiebung der Masse. Die Variable F_0 beschreibt die Kraftamplitude und Ω die konstante Kreisfrequenz/Winkelgeschwindigkeit. Dadurch ergibt sich der Zusammenhang zum Winkel $\varphi = \Omega t$. [1], [2]

Der Untergrund kann als Mehrfreiheitsgradsystem (MDOF) modelliert werden. Wird dieses Modell in den *Modalen Raum* überführt, ergibt sich ein entkoppeltes System aus Einmassenschwingern (SDOF). Dabei beschreibt jeder Einmassenschwinger einen Freiheitsgrad bzw. eine Eigenfrequenz des Mehrfreiheitsgradsystems (Abb.2).

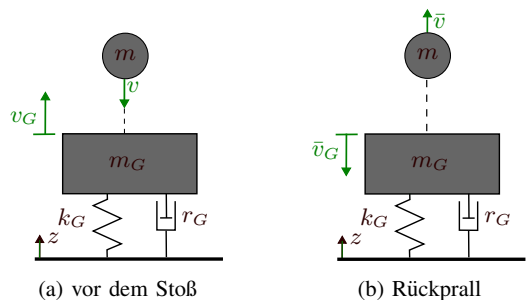


Abb. 2: Zentraler Stoß. a) vor dem Stoß, b) Rückprall.

Aktuell wird eine Masse $m = 62,5 \text{ kg}$ aus einer Höhe $h = 1,2 \text{ m}$ auf den Untergrund fallen gelassen. Diese Masse m führt einen teilelastischen Stoß aus (Abb.2). Die Masse prallt wieder vom Untergrund (m_G, k_G und r_G) zurück. Wird der erneute Fall nicht gebremst, indiziert die Masse einen Mehrfachimpuls in den Untergrund.

Um die auftretende Kraft beim Aufprall der Masse m auf den Untergrund zu bestimmen, wird die

Impulserhaltung benötigt. Unter der Annahme einer zeitunabhängigen Masse gilt:

$$m(v + \bar{v}) = \int_0^{t_s} F_p(t) dt. \quad (2)$$

Dabei beschreiben die Variablen v und \bar{v} die Geschwindigkeiten unmittelbar vor und nach dem Stoß. Weiters beschreibt die Variable $F_p(t)$ den Kraftverlauf über die Zeit und die Variable t_s die Stoßzeit.

Der Zusammenhang der Geschwindigkeiten vor und nach dem Stoß bzw. der Höhen vor und nach dem Stoß kann über die Stoßzahl, unter Vernachlässigung der Untergrundgeschwindigkeiten v_G und \bar{v}_G , berechnet werden zu:

$$e_s = \frac{\bar{v}}{v} = \frac{\sqrt{2gh}}{\sqrt{2gh}} = \sqrt{\frac{\bar{h}}{h}}. \quad (3)$$

Die Variablen v und \bar{v} beschreiben die Geschwindigkeiten vor und nach dem Stoß und die Variablen h und \bar{h} beschreiben die Höhen vor und nach dem Stoß. [3]

Ziel dieser Arbeit ist die Durchführung der Experimentellen Modalanalyse an Schienensystemen. Dabei gilt es die Eigenfrequenzen der anzuregenden Struktur und die Übertragungsfunktion zwischen Schienen und Nachbargebäuden zu bestimmen. Die Übertragungsfunktion beschreibt den Zusammenhang zwischen den Größen am Ausgang (Schwingungsantwort des Systems) und Eingang (Anregung des Systems) eines schwingfähigen Systems. Für die Berechnung der Übertragungsfunktion im Kontinuierlichen wird Gleichung (1) in den *LAPLACE Bereich*

$$(ms^2 + rs + k)Q(s) = F(s) \quad (4)$$

überführt und so die Transferfunktion

$$G(s) = \frac{Q(s)}{F(s)} = \frac{1}{ms^2 + rs + k} \quad (5)$$

gebildet. Die Variable $Q(s)$ beschreibt die Schwingungsantwort des Systems und die Variable $F(s)$ die Anregung des Systems. Für die Experimentelle Modalanalyse wird jedoch die komplexe Frequenzdarstellung bevorzugt. Durch das Einsetzen von $s = i\Omega$ in Gleichung 5 ergibt sich die Frequenzdarstellung der Übertragungsfunktion zu:

$$H(i\Omega) = \frac{Q(i\Omega)}{F(i\Omega)} = \frac{1}{-m\Omega^2 + ri\Omega + k}. \quad (6)$$

Die Signale am Ausgang und Eingang eines schwingfähigen Systems werden zeitsynchron aufgenommen und mittels FFT in den Frequenzbereich transformiert. Diese diskretisierten Signale liegen dann als komplexe Spektren vor. Danach wird die komplexe Übertragungsfunktion gebildet. Das bedeutet, es gibt für jede Frequenz eine Amplitude (Realteil) und einen Phasenwinkel (Imaginärteil). Normalerweise werden Amplituden- und Phasenfrequenzgang über die Kreisfrequenz Ω oder die Frequenz f aufgetragen. Durch Berechnen der Pole der Übertragungsfunktion in der komplexen Ebene ergeben sich die Eigenfrequenzen des schwingfähigen Systems. Die Bestimmung der Übertragungsfunktion bei einer Kraftanregung mit Mehrfachimpulsen wird jedoch erheblich erschwert. [4]

Um den Referenzimpuls I_{ref} (Masse $m_{ref} = 62,5$ kg, Fallhöhe $h_{ref} = 1,2$ m) zu erreichen, wird für den Impulshammer eine Fallmasse mit $m = 45$ kg verwendet. Um den selben Impuls zu erzielen, muss aufgrund der Massenreduktion die Fallhöhe h erhöht werden. Aus dem Referenzimpuls I_{ref} wird die benötigte Aufprallgeschwindigkeit, unter Berücksichtigung der Erdbeschleunigung $g = 9,81$ m/s², der Masse auf den Untergrund

$$v = \frac{I_{ref}}{m} = \frac{I_{ref}}{\sqrt{2gh_{ref}}} = 6,80 \text{ m/s} \quad (7)$$

berechnet. Daraus kann anschließend die benötigte

Fallhöhe

$$v = \sqrt{2gh} \rightarrow h = \frac{v^2}{2g} = 2,315 \text{ m} \quad (8)$$

berechnet werden.

Der entwickelte Impulshammer (Abb.3) kann auf die anzuregende Struktur gestellt werden. Er wird senkrecht über die einzustellenden FüÙe ausgerichtet und anschließend kann bereits mit der Anregung der Struktur begonnen werden. Dafür wird die Masse über eine Seilwinde auf die vorgegebene Fallhöhe gebracht und gelöst.

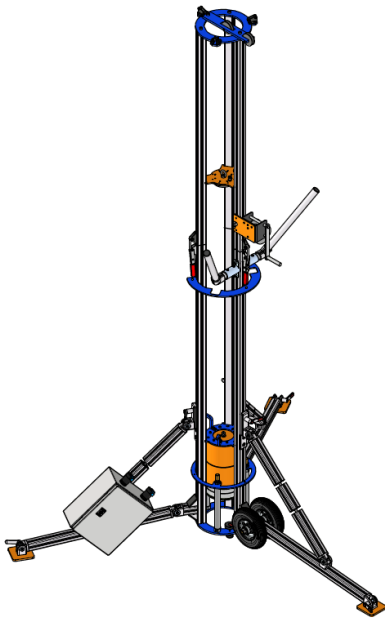


Abb. 3: Impulshammer im Einsatz

Für das sichere Halten bzw. Lösen der Masse wird ein Nullpunktspannsystem verwendet. Die Masse prallt auf den Untergrund auf und indiziert eine Anregungskraft. Der Kraftverlauf $F_p(t)$ wird über

einen für Stoßanwendungen entwickelten Kraftsensor (DMS basiert) gemessen [5]. Dieser Sensor ist am Boden der Fallmasse montiert und liefert ein driftfreies Spannungssignal. Danach prallt die Masse vom Untergrund zurück und wird von der Bremsvorrichtung (Abb.4) zum Stillstand gebracht.

Durch das Abbremsen der Masse kann der Mehrfachimpuls verhindert bzw. reduziert werden. Für den Impulshammer wird daher eine Bremsvorrichtung vorgesehen (Abb.4). Diese Vorrichtung verhindert das erneute Auftreffen der Masse auf den Untergrund nach dem Rückprallen des ersten Stoßes. Dabei wird beim Bremsvorgang darauf geachtet, dass die Hubzylinder (Pos. 200.70) die Masse m vor erneutem Aufprall auf den Untergrund abbremsen. Die Spitze (Pos. 3.1) berührt den Untergrund beim Bremsvorgang nicht. Die Masse wird über die Spitzen der Hubzylinder (Pos. 1.5) und Lasche der Fallmasse (Pos. 3.9) abgebrems/ aufgefangen.

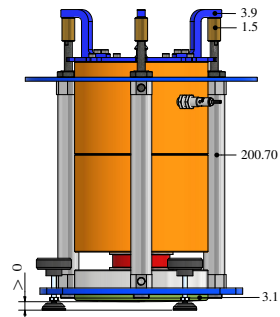


Abb. 4: Bremsvorrichtung der Masse

Die Bremsvorrichtung wird pneumatisch ausgeführt. Das hat den Vorteil, dass einerseits die Bremskraft über den pneumatischen Druck und andererseits die Bremsgeschwindigkeit über eine Abluftdrosselung eingestellt werden kann. So kann der Bremsimpuls über einen möglichst langen Bremsweg der Hubzylinder mit geringer Amplitude abgebaut werden. Das hat den Vorteil, dass die überlagerte Anregung durch die Mehrfachstöße minimiert wird.

Ist der Versuch beendet, kann die Konstruktion in die Transportstellung (Abb.5) gebracht werden. Dafür wird der obere Teil des Impulshammers in den unteren Teil gesteckt und die Masse über die Seilwinde an die Profile gezogen. Der Impulshammer kann dann über die Räder zur nächsten Messstelle bewegt werden.

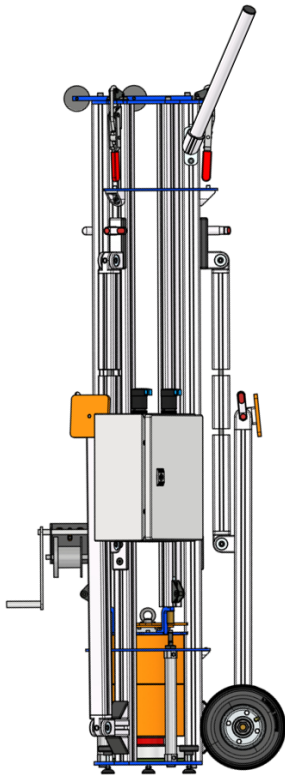


Abb. 5: Impulshammer in Transportstellung.

III. ERGEBNISSE

Hier werden die Messdaten der Impulshammer (Ausgangssituation/ Neuentwicklung) verglichen. Die Messdaten der angeregten Struktur werden nicht näher behandelt. Um einen Eindruck der Problemstellung mit den Mehrfachimpulsen

zu bekommen, wird das gemessene Resultat des Impulshammers mit der Ausgangssituation verglichen.

Abbildung 6 zeigt den Kraftverlauf $F_p(t)$, der Ausgangssituation, im Zeitbereich. Es sind deutlich die Mehrfachstöße erkennbar. Ersichtlich sind die Messdaten ($F_{measure}$) und die bereinigten Messdaten (F_{modify}). Die nachfolgenden Berechnungen werden mit den bereinigten Daten gemacht.

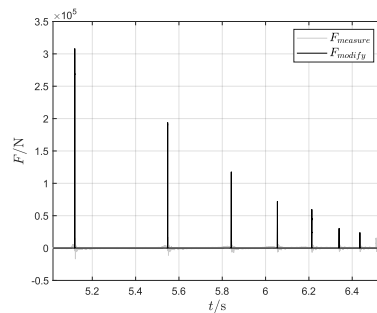


Abb. 6: Kraftverlauf (Ausgangssituation freier Fall) $F_p(t)$ im Zeitbereich.

Diese Mehrfachstöße im Zeitbereich verursachen ein überlagertes Signal im Frequenzbereich (Abb.7).

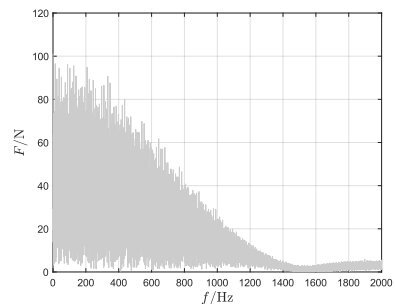


Abb. 7: FFT Kraftverlauf Ausgangssituation.

Durch dieses überlagerte Signal ist eine eindeutige Bestimmung der Übertragungsfunktion zwischen

OUTPUT (gemessene Beschleunigung der angeregten Struktur) und INPUT (Kraftverlauf des Impulshammers) nicht möglich.

Abbildung 8 zeigt den Kraftverlauf $F_p(t)$ des neu entwickelten Impulshammers im Zeitbereich. Es sind keine Mehrfachstöße mehr erkennbar. Jedoch ist der Bremskraftverlauf leicht erkennbar.

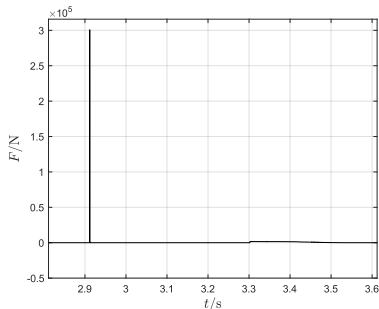


Abb. 8: Kraftverlauf Impulshammer $F_p(t)$ im Zeitbereich.

Durch das Verhindern der Mehrfachstöße bei der Kraftanregung erhält man eine weniger überlagerte Kraftanregung und daraus ein weniger überlagertes Signal im Frequenzbereich (Abb.9).

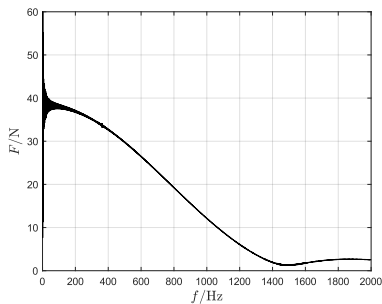


Abb. 9: FFT Kraftverlauf Impulshammer.

Die Überlagerung am Beginn des Signals ist auf das Lösen der Masse und den Bremsvorgang

zurückzuführen. Mit solch einem Signal ist eine eindeutige Bestimmung der Übertragungsfunktion möglich.

IV. CONCLUSIO

Wie aus den Ergebnissen ersichtlich ist, kann durch das Abbremsen der Masse ein erneuter Aufprall auf den Untergrund verhindert werden. Durch das Verhindern der Mehrfachstöße wird eine deutliche Reduktion der Überlagerung im Frequenzbereich erzielt. Dadurch kann eine eindeutige Übertragungsfunktion bestimmt werden. Weiters ist es gelungen, die Masse der Vorrichtung zu reduzieren und einen einfacheren Transport zu ermöglichen.

LITERATUR

- [1] F.J. Falkner, "Lecture notes machine dynamics," Skriptum, 2022, Management Center Innsbruck.
- [2] L.Moschen, "Vorlesungsunterlagen technische mechanik 3," Skriptum, 2019, Management Center Innsbruck.
- [3] D. Gross, W. Hauger, J. Schröder, W. A. Wall, *Technische Mechanik 3 (Kinetik)*, 15th ed. Springer, 2021.
- [4] T. Kuttner, *Praxiswissen Schwingungsmesstechnik*. Springer, 2015.
- [5] Hottinger Brüel und Kjaer, *Bedienungsanleitung KMR+ Mat.: 7-0111.0006 DVS: A05429_01_YCI_02 HBM: public 11.2020*, Hottinger Brüel und Kjaer, accessed: 2023-06-10.



Michael Kofler arbeitet bei der Firma INNIO Jenbacher GmbH & Co. OG in der Arbeitsvorbereitung für die mechanischen Fertigung. Im Zuge seines Masterstudiums am MCI befasste er sich mit der Entwicklung eines Impulshammers für Schienensysteme.

Intelligenter Entmistungsroboter für die Reinigung von Spaltböden

Thomas Lederer und Ronald Stärz (Betreuer)

Kurzfassung—Die vorliegende Arbeit befasst sich mit der Konzeptionierung, Konstruktion, elektrischen Auslegung und dem Prototypenbau eines vollautomatisierten Systems zur Reinigung von Spaltenböden in Kuhställen. Da ein immer größerer Teil der Kühe nicht mehr in Anbinde- sondern in Laufställen gehalten wird, in welchen Spaltenböden als Laufflächen fungieren, ist eine wiederkehrende Reinigung dieser unerlässlich. Derzeit verfügbare Systeme sind durch zahlreiche Adaptionsarbeiten in den Ställen mit einem erheblichen Kostenaufwand verbunden, wodurch diese für den "kleinen" Landwirtschaftsbetrieb unerschwinglich sind.

In der Arbeit wird die mechanische Auslegung sämtlicher Komponenten des Spaltenroboters behandelt. Hierzu werden sämtliche mögliche Belastungen ermittelt und in eine analytische Berechnung miteinbezogen, um die entsprechenden Maße zu erhalten, bevor diese für die Konstruktion in der 3D-Software INVENTOR verwendet werden. Außerdem werden alle notwendigen Fertigungszeichnungen erstellt. Anschließend werden sämtliche elektrische Komponenten definiert, was die Dimension der notwendigen Speicherkapazität der verwendeten Akkus sowie die Berechnung der benötigten Leistung der verwendeten Motoren beinhaltet. Zudem beinhaltet die elektrische Auslegung die grundlegende Konzeption einer Trägerplatine, welche den verwendeten Mikrocontroller mit der benötigten Sensorik und etwaigen Steuerungselementen sowie der Spannungsversorgung verbindet. Bezüglich der Sensorik werden verschiedenste, für den mobilen Roboter geeignete Sensortypen ermittelt und miteinander verglichen. Hierbei werden vor allem

Ultraschallsensoren und optische Sensorsysteme, wie Tiefenkameras und 3D-Laserscanner, betrachtet. Die entworfene Softwarearchitektur für die Navigation des Roboters durch das Stallgebäude stützt sich auf eine hybride Architektur zwischen einer verhaltens- und funktionsorientierten Programmarchitektur, um sämtliche Vorteile miteinander zu vereinen. Zudem werden simultane Positions- und Kartieralgorithmen ermittelt, welche verwendet werden, um eine Karte der lokalen Umgebung - des Stalles - zu erstellen und die exakte Position des Roboters zu ermitteln.

Abschließend wird mit der Fertigung eines Prototypen mit den gewonnen Erkenntnissen begonnen. Dabei wird das mechanische Design realisiert sowie die elektrischen Bauteile im Roboter platziert. Die Anbringung der Sensorik und Programmierung des Mikrocontrollers wird zu einem späteren Zeitpunkt vollendet, da aufgrund von finanziellen Engpässen und Lieferschwierigkeiten nicht alle Bauteile verfügbar sind.

Schlagwörter—Reinigung von Spaltenböden; Spaltenroboter; Vollautomatisiertes System; Kartierungs- und Positionierungsalgorithmen; Prototypenbau

I. EINLEITUNG

EIN immer größerer Teil der Kühe wird nicht mehr in Anbinde- sondern in Laufställen gehalten. Dadurch wird das Wohlbefinden der Kühe durch ein sehr großräumiges Bewegungsangebot gefördert. Diesem Vorteil steht allerdings der Umstand entgegen, dass vor allem die Klauen der Tiere einer hohen Belastung durch Schmutz auf der Lauffläche ausgesetzt sind. Da in einem Laufstall die Spaltenböden als Laufflächen fungieren, ist eine wiederkehrende Reinigung von diesen unerlässlich, um eine Vorbeugung gegen diverse Krankheiten zu gewährleisten.

T. Lederer studiert am Studiengang Mechatronik, MCI, Innsbruck, Österreich, e-mail: lt5481@mci4me.at.

R. Stärz arbeitet am Studiengang Mechatronik, MCI, Innsbruck, Österreich.

Manuskript eingereicht am 29. März 2022; revidiert am 31. März 2022.

Abbildung 1 zeigt hierbei die Haltungsform in einem Laufstall.



Abbildung 1. Laufstallhaltung von Kühen

Zudem hängt die Leistungsbereitschaft einer Kuh nicht lediglich von der Genetik und den Futterrationen, sondern zu einem maßgeblichen Teil auch von der Hygiene der Stallgegebenheiten ab, was die Notwendigkeit einer regelmäßigen Reinigung abermals unterstreicht. Da die Produktion höchster Qualitätsstandards im Fokus eines jeden Landwirtes steht und diese Hand in Hand mit einer hygienischen Umgebung für die Tiere einhergeht, soll abermals die Unausweichlichkeit einer wiederholten Reinigung untermauert werden.

Um gewünschte Resultate zu erzielen, muss die Reinigung der Spaltenböden in der Praxis mindestens zweimal täglich vollzogen werden. Abbildung 2 zeigt hierbei den ungesäuberten im Vergleich zum gesäuberten Spaltenboden.

Die Arbeit der Spaltenböden ist hierbei sehr monoton und mit einem erheblichen Zeitaufwand verbunden. Weiters birgt die Arbeit ebenso potentielle Gefahren, da ein eventuell mit den Kühen mitlaufender Stier oder ein Muttertier den Arbeiter als Feind sehen und diesen deshalb attackieren könnte. Diese Umstände betrachtet, ist die Automatisierung der Reinigung der Spaltenböden ein naheliegender Ansatz. Aktuell verfügbare Systeme haben allerdings oft den Nachteil, dass erhebliche Modifikationen des Stalles notwendig sind, um



Abbildung 2. Ungesäuberter und gesäuberter Spaltenboden im Vergleich

die Funktion zu gewährleisten, was sich in den Investitionskosten niederschlägt.

Die vorliegende Arbeit soll dazu dienen, ein voll automatisiertes System zur Reinigung von Spaltenböden in Laufställen, in Form eines Entmistungsroboters zu entwerfen. Hierbei wird sowohl die mechanische Auslegung, die Auswahl elektronischer Komponenten, sowie die Konzipierung und Entwicklung der benötigten Software behandelt. Zudem wird ein entsprechender Prototyp gebaut, um die korrekte Funktionsweise zu evaluieren.

II. SYSTEMSPEZIFIKATION UND SYSTEMDESIGN

Der folgende Abschnitt befasst sich hauptsächlich mit der Grundidee für die Realisierung des Systems. Darüber hinaus werden die Anforderungen und Aufgaben, welche erfüllt werden müssen, genau erläutert.

Abbildung 3 zeigt den prinzipiellen Aufbau des Systems mit den grundlegenden Komponenten und Gegebenheiten, mit welchen geeignet umgegangen werden muss, um einen ordnungsgemäßen Betrieb zu

gewährleisten. Wie dieser Abbildung entnommen werden kann, stellen die Kühe dynamische Hindernisse für das System dar, weshalb eine geeignete Programmstruktur zur Reaktion auf diese entworfen werden muss.

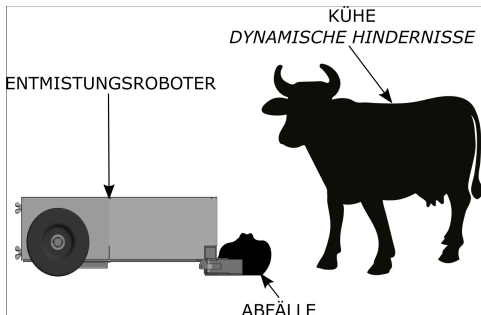


Abbildung 3. Systemspezifikation

Aus Abbildung 3 lassen sich bereits die generellen Anforderungen an das automatisierte Reinigungskonzept in Form eines Roboters bestimmen. Diese sind in Tabelle I aufgelistet.

Tabelle I
GESAMTANFORDERUNGEN AN DEN ROBOTER

Anforderung	Vorgabe
Maximale Gewichtsbelastung	1300 kg
Auflagegewicht Reinigungsschiene auf Spaltenboden	min. 20 kg
Maximale Abmaße (Länge x Breite x Höhe)	120 x 120 x 42 cm
Schubleistung	min. 35 kg
Minimale Räumbreite	min. 80 cm
Minimale Räumgeschwindigkeit	4 m/min
Maximale Räumgeschwindigkeit	15 m/min
Maximale Räumdauer	3 h

Tabelle II listet ebenso die Hauptanforderungen an die Ladestation für den Roboter auf, welche aus den baulichen Gegebenheiten der Stallanlage resultieren.

Tabelle II
GESAMTANFORDERUNGEN AN DIE LADESTATION

Anforderung	Vorgabe
Maximale Anschlussleistung	5 kW
Maximale Abmaße (Länge x Breite x Höhe)	120 x 120 x 42 cm
Abdichtung Gehäuse	Spritzwasserdicht

A. Mechanische Auslegung

Für die mechanische Auslegung des Spaltenroboters werden sämtliche bekannte Belastungen ermittelt und in eine analytische Berechnung miteinbezogen, um die entsprechenden Maße für die anschließende Konstruktion in der 3D-Software INVENTOR zu erhalten.

Die für die mechanische Berechnung notwendigen Belastungen können hierbei größtenteils Tabelle I entnommen werden, welche im Folgenden aufgelistet sind:

- Maximale Gewichtsbelastung: 1300 kg
Diese Angabe basiert auf dem möglichen Gewicht eines mitlaufenden Stieres
- Auflagegewicht Reinigungsschiene: min. 20 kg
- Schubleistung: min. 35 kg

Das Herzstück des mechanischen Antriebes des Entmistungsroboters sind die Antriebsachsen der beiden Gummiräder, welche im Vollschnitt in Abbildung 4 dargestellt ist.

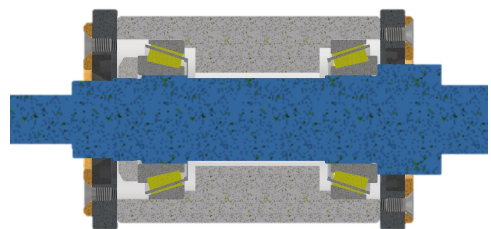


Abbildung 4. Angestellte Lagerung der Antriebsachse

Wie in Abbildung 4 ersichtlich ist, wird eine angestellte Lagerung mit den einreihigen

Kegelrollenlagern 320/32 X von SKF [1] in O-Anordnung angewendet. Da bei der Verwendung der O-Anordnung die Drucklinien nach außen zeigen und somit die Stützbasis größer ist, ergibt sich eine starre Lagerung, welche bei Wärmedifferenzen weniger zum Verspannen der Ringe neigt [2].

Abbildung 5 zeigt das Grundgerüst des Roboters, welches aus geschweißten Profilen besteht. Zudem ist der Zwischenboden ersichtlich, welcher im weiteren Projektverlauf zur Unterbringung der Steuerungselektronik und der Motortreiber verwendet werden soll.

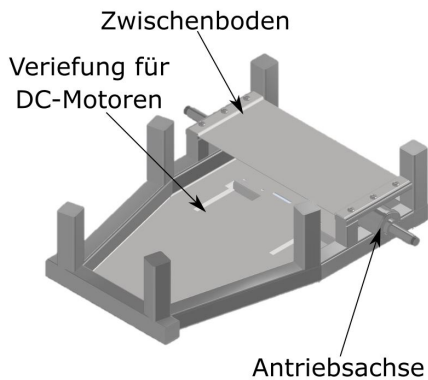


Abbildung 5. Profilrahmen

Abbildung 5 zeigt weiters die zuvor beschriebenen Antriebsachsen sowie Vertiefungen für die verwendeten DC-Motoren.

In Abbildung 6 ist der geschlossene und fertig zusammengebaute Roboter ersichtlich. Die verwendeten 10 Zoll Reifen weisen hierbei ein Stapler-Blockprofil auf und haben eine 2-teilige Stahlfelge mit einer Antriebsnabe für eine $8 \cdot 3 \text{ mm}$ Keilnut.

Auf der Vorderseite ist der Mistschieber, welcher eine Gummilippe aufklebt hat, um den Spaltenboden zu schonen und eine saubere Reinigung zu gewährleisten.

welcher eine Gummilippe aufklebt hat, um den Spaltenboden zu schonen und eine saubere Reinigung zu gewährleisten.

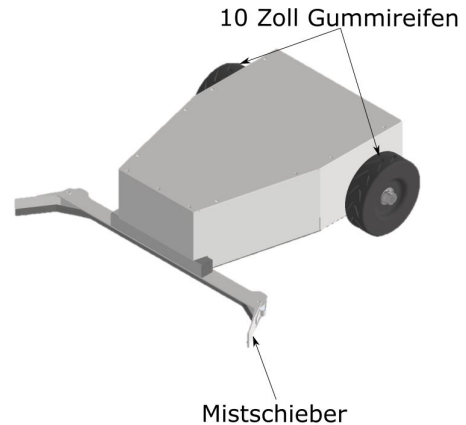


Abbildung 6. Gesamtzusammenbau Roboter

B. Elektronische Auslegung

Die für die elektronische Auslegung notwendigen Größen können ebenso größtenteils Tabelle I entnommen werden und sind im Folgenden erneut aufgelistet:

- Schubleistung: min. 35 kg
- Räumgeschwindigkeit: $4 - 15 \frac{\text{m}}{\text{min}}$
- Maximale Räumdauer: 3 h

Aus der minimalen Schubleistung von 35 kg kann das notwendige Drehmoment der Motoren ermittelt werden. Da die Schubleistung auf zwei Motoren aufgeteilt wird, muss ein Motor lediglich eine Schubleistung von 17.5 kg bereitstellen, welche laut Gleichung 1 einer Kraft von 171.68 N entspricht.

$$F_S = m g$$

$$= 17,5 \cdot 9,81 \frac{\text{kg} \cdot \text{m}}{\text{s}^2} = 171,68 \text{ N} \quad (1)$$

Durch das Gleichsetzen des Motordrehmomentes mit dieser Kraft und dem Durchmesser der Räder von 10 Zoll, kann das notwendige Drehmoment nach Gleichung 2 ermittelt werden.

$$M_M = F_S \frac{D_R}{2} \quad (2)$$

$$M_M = 171,68 \cdot \frac{0,254}{2} Nm = 21,8 Nm$$

Um ein sicheres Schieben zu gewährleisten und etwaige Überbelastungen der Motoren zu vermeiden, wird ein Sicherheitsfaktor von 2 angewendet, woraus sich ein Drehmoment von $43,6 Nm$ ergibt. Um einen geringen Wartungsaufwand der Motoren zu gewährleisten, werden bürstenlose DC-Motoren mit einem Schneckengetriebe verwendet, welche ein Drehmoment von $46 Nm$ bereitstellen. Die Ansteuerung dieser Motoren wird mithilfe des Motortreibers *ATO BLD-50A* vollzogen, welcher im Lieferumfang der entsprechenden Motoren enthalten ist.

Da die verwendeten Motoren einen Nennstrom von je $20,8 A$ benötigen und die maximale Räumdauer $3 h$ beträgt, ergibt sich laut Gleichung 3 die notwendige Akkukapazität des Roboters.

$$Q_{Robo} = I_{Rated} T_{Räum} n_m \quad (3)$$

$$Q_{Robo} = 20,8 \cdot 3 \cdot 2 Ah = 124,8 Ah$$

Da der BLDC-Controller [3] des DC-Motors laut Datenblatt am Spannungseingang einen typischen Wert von $36 V$ erwartet, wird die notwendige Kapazität mithilfe von drei seriell verschalteten $12 V$ Akkus aufgebracht. Hierfür werden drei wartungsfreie *BTL 12-55L* Blei-Vlies-Akkumulatoren von *Effekta* [4] mit einer Spannung von $12 V$ und einer Kapazität von $55 Ah$ verwendet. Zum Laden der in Serie geschalteten Akkumulatoren wird das $36 V$ Ladegerät *Mascot 2047 Bleilader* verwendet. Durch den Einsatz von lediglich einem Ladegerät können die Ladeanschlüsse sowie auch die Ladestation einfacher gestaltet werden als

beispielsweise bei der Verwendung von drei $12 V$ Ladegeräten.

Abbildung 7 zeigt den beschriebenen Zusammenbau mit den drei Akkumulatoren und den zwei bürstenlosen DC Motoren.

Für die grundlegende Konzeption der Trägerplatte werden sämtliche Verbindungen vom MCU zu den Motortreibern oder etwaigen anderen elektrischen Komponenten mithilfe von Optokopplern realisiert, wie in Abbildung 8 schematisch ersichtlich ist.

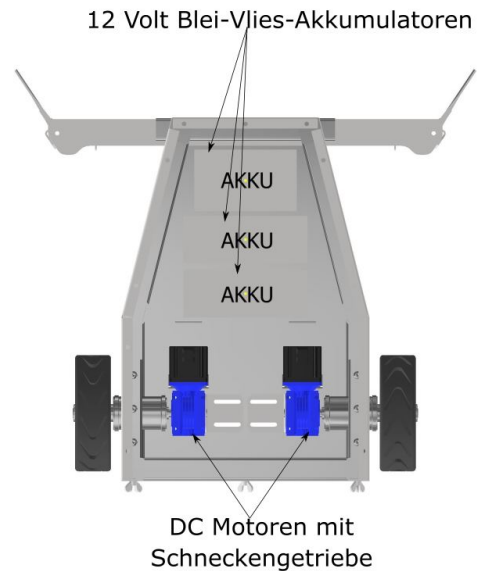


Abbildung 7. Akkus und Motoren im Zusammenbau

Durch die Verwendung der Optokoppler für die Schnittstelle zwischen dem Mikrocontroller und den elektronischen Steuerungselementen wird eine galvanische Trennung der entsprechenden Netze erreicht. Somit wird dieser vor eventuellen Beschädigungen durch etwaige Rückkopplungen und Fremdeinflüssen geschützt.

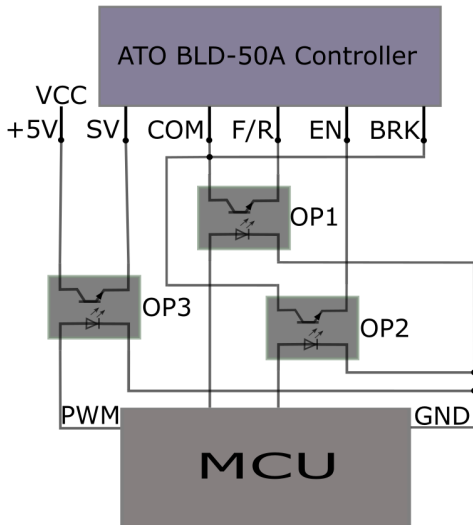


Abbildung 8. Schematische Darstellung der Ansteuerung des Motortreibers über die MCU

Für die Umsetzung des Projektes werden zudem verschiedene Sensoriksysteme zur Erfassung der dynamischen Umgebung des Stalles miteinander verglichen, um den am besten geeigneten Sensortyp zu ermitteln. Die Rinder stellen hierbei bewegte und ortsunabhängige Hindernisse dar, welche während des Reinigungsprozesses ständig evaluiert werden müssen. Besonderes Augenmerk zur Sensorauswahl wird auf den finanziellen Aspekt und auf die Zuverlässigkeit der Messungen, wie auch auf die Messperformance gelegt. Hierzu werden **Ultraschallsensoren** und die optischen Erfassungssysteme **LiDAR** sowie **Tiefenkameras** miteinander verglichen.

Alle drei Sensoriksysteme basieren hierbei auf dem Prinzip der Laufzeitmessung, welches schematisch in Abbildung 9 dargestellt ist.

Hierbei wird prinzipiell die benötigte Zeit zum Objekt und wieder zurück der ausgesendeten

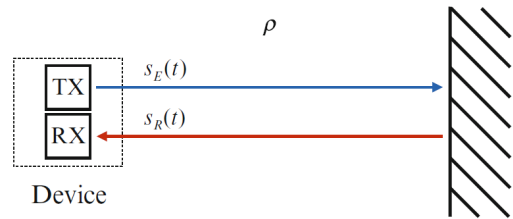


Abbildung 9. Schematische Darstellung des Prinzips der Laufzeitmessung [5]

Schallwellen beziehungsweise der Licht- oder Laserstrahlen gemessen. Mithilfe der bekannten Geschwindigkeit des ausgesendeten Mediums kann anschließend die entsprechende Distanz ermittelt werden. [6]

Verglichen mit den optischen Erfassungssystemen bietet die Ultraschallsensorik in der Stallumgebung den Vorteil, dass diese Systeme sehr robust und gegenüber Staub- sowie Feuchtigkeitseinflüssen resistent sind. Zudem sind Ultraschallsensoren in den verschiedensten Anwendungsgebieten vertreten, wodurch verschiedenste Ausführungen von diesen Sensortypen verfügbar sind. Auf Basis dieser Gegebenheiten und Anforderungen hat sich der Ultraschallsensor *Pepperl+Fuchs UB4000-F42-I-V15* für den weiteren Projektverlauf als passendes Erfassungssystem mit einer Reichweite von 4000 mm und einer Blindzone von lediglich 20 mm abgezeichnet [7].

Die vorgestellten optischen Erfassungssysteme in Form der Tiefenkamera *RealSense D435i* [8] sowie dem 3D-Laserscanner *RS-BPearl 360* [9] bieten hingegen den Vorteil, dass ein deutlich größerer Bereich observiert werden kann, wodurch eine vorausschaulichere Umsetzung der Routenplanung und Navigation möglich ist. Da die Notwendigkeit einer robusten Ausführung des Sensorgehäuses aufgrund der rauen Umgebung gegeben ist, ist

die Verwendung der Tiefenkamera nicht möglich. Darum hat sich in Hinsicht auf die optischen Erfassungssysteme der 3D-Laserscanner *RS-BPearl 360* für die Umsetzung durchgesetzt.

Zur Realisierung des vorliegenden Systems wird auf Basis obiger Argumentation und unter Berücksichtigung der Kosten abhängig von der finanziellen Situation des Autors entweder die Ultraschallsensorik mit den Sensoren *Pepperl+Fuchs UB4000-F42-I-V15* oder das LiDAR System mit dem 3D-Laserscanner *RS-BPearl 360* verwendet.

C. Auslegung der Softwarearchitektur

Bei der Auslegung der Softwarearchitektur wird im ersten Schritt der geeignete Mikrocontroller für die Umsetzung der entsprechenden Software ausgewählt. Dabei wird sichergestellt, dass dieser die notwendige Rechenleistung und Speicherkapazität zur Verfügung stellt. Zur Umsetzung der ausgearbeiteten Software wird darum das Evaluationskit *Atmel SAM V71 Xplained Ultra* [10] ausgewählt, welches eine enorme Rechenleistung wie auch zahlreiche Schnittstellen zur Verfügung stellt.

Hinsichtlich der Routenplanung und Kartenerstellung unterscheidet sich das vorliegende Projekt in Hinsicht auf andere Anwendungsgebiete für autonome, mobile Roboter dadurch, dass die Umgebungsgrenzen bereits im Vorhinein durch die Stallwände vollständig definiert und überschaubar sind. Die idealen Routen des Spaltenroboters sind hierbei in Abbildung 10 ersichtlich, welche aus dem zu reinigenden Arbeitsbereich und der Dimension des Mistschiebers resultieren. Die obig skizzierten Routen können allerdings durch die dynamischen Hindernisse, welche in Form von Rindern auftreten, beeinflusst werden.

Die verhaltensbasierte Softwarearchitektur basiert auf einem reinen Reaktionssystem. Prinzipiell werden hierbei also lediglich die Eingangswerte der Sensoren zur Laufzeit verwendet, um Entscheidungen zu

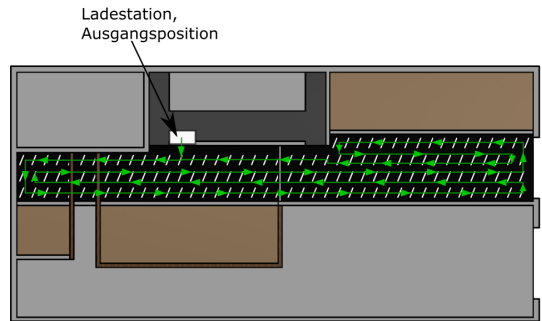


Abbildung 10. Ideale Reinigungsroute

treffen [11]. Dies bietet den Vorteil, dass zum einen nur sehr geringe Rechenkapazitäten notwendig sind und zum anderen extrem schnell auf dynamische Hindernisse reagiert werden kann. Allerdings kann es durch diese Struktur vorkommen, dass Fehler wiederholt werden und im schlimmsten Fall in einer Endlosschleife enden, wie in Abbildung 11 dargestellt ist.

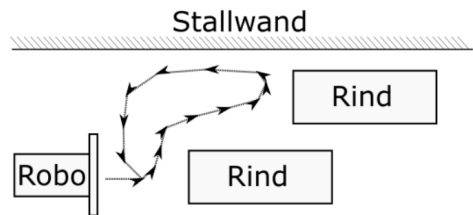


Abbildung 11. Mögliche negative Folgen eines verhaltensbasierten Ansatzes

Die funktionsorientierte Softwarearchitektur basiert laut [12] auf einem deliberativen Ansatz, welcher vorerst ein Weltbild in Form einer Karte aufbaut. Durch eben diese Kartierung ist es mit dem funktionsorientierten Ansatz möglich, die Navigation des Spaltenroboters durch die zu reinigende Umgebung vorausschauend zu planen. Zudem zeichnet sich dieser Ansatz vor allem durch die feste Abfolge von vordefinierten Zuständen aus.

Für die Planung der Navigation fließen dabei sowohl die momentane Situation durch Sensormessungen und die interne Repräsentation der Umgebung mitein [13]. Für die Erstellung des Welbildes, beispielsweise mithilfe von SLAM-Algorithmen, bedarf es allerdings einen leistungsstarken Mikrocontroller. Zudem kann durch den hierarchischen Aufbau dieser Struktur nicht schnell auf dynamische Hindernisse reagiert werden, da hierzu vorerst der bereits veraltete Prozess abgeschlossen und neu durchlaufen werden muss.

Basierend auf den Erkenntnissen der vorgestellten Architekturen kann zusammenfassend festgestellt werden, dass sowohl die verhaltensbasierte wie auch die funktionsorientierte Architektur Eigenschaften mit sich bringen, welche für eine problemlose Prozessdurchführung unerlässlich sind. Während der hierarchische Ansatz den Vorteil der genaueren und vorausschauenderen Planung sowie der exakten Positionsbestimmung bietet, so stellt der reaktive Ansatz die spontane Reaktion auf sich bewegende Hindernisse zur Verfügung. Darum wird eine hybride Softwarearchitektur erarbeitet, welche die Vorzüge beider Architekturen miteinander vereint. Prinzipiell basiert die erarbeitete hybride Struktur auf dem verhaltensbasierten Ansatz. Somit wird erreicht, dass die Spontanität hinsichtlich der Reaktion auf dynamische Hindernisse erhalten bleibt. Allerdings wird in diesem Ansatz die Planungsphase des hierarchischen Ansatzes miteinbezogen, um auch dessen Vorteile mitzubringen, wie in Abbildung 12 ersichtlich.

Die Sensordaten werden hierbei ohne aufwändige Verarbeitung direkt auf die auszuführenden Aktionen, welche sich in obigem Flussdiagramm auf die korrekte Beschaltung der Motortreiber beschränken, abgebildet. Zudem werden die ermittelten Sensorinformationen allerdings an eine Planungskomponente weitergegeben, welche die entsprechenden Aktionen zudem im Sinne einer vorausschauenden Planung beeinflussen.

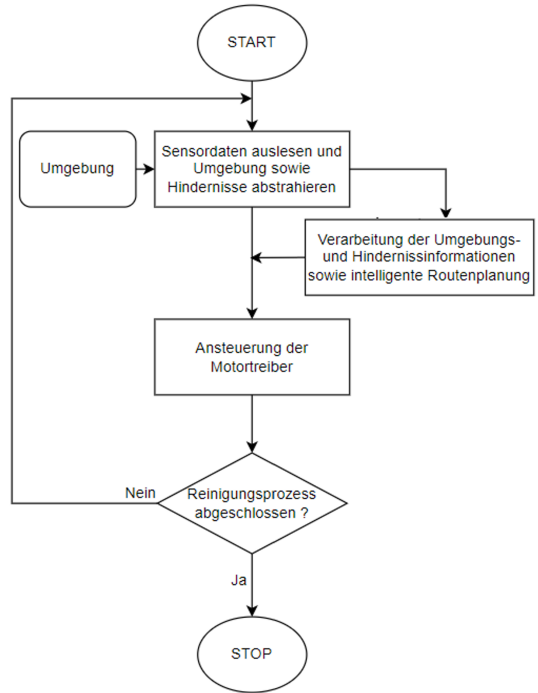


Abbildung 12. Flussdiagramm der hybriden Softwarearchitektur

III. REALISIERUNG

Mithilfe der gewonnen Erkenntnisse wird die Anfertigung und Umsetzung eines Prototypen vollzogen, soweit dies die finanzielle Situation des Autors und die Lieferbedingungen zulassen. Hierbei wird vorerst das Grundgerüst mithilfe der angefertigten Konstruktionszeichnungen aus oxidationsbeständigem Stahl realisiert und die DC Motoren während des Zusammenbaus an den Antriebsachsen montiert. Anschließend werden die Akkumulatoren sowie die Steuerungselemente wie die Motortreiber platziert. Der Stand des Prototypen zum momentanen Zeitpunkt ist in Abbildung 13 abgebildet.



Abbildung 13. Prototyp des Spaltenroboters ohne Deckel

IV. CONCLUSIO

Die vorliegende Arbeit bietet einen Einblick in die Entwicklung eines autonomen mobilen Spaltenroboters. Zudem wird mit der Umsetzung eines Prototypen auf Basis der gewonnenen Kenntnisse begonnen. Die Entwicklung eines solchen Robotersystems ist allerdings mit einem enormen Zeitaufwand verbunden und bedarf einem fundiertem Wissen im Bereich der verschiedenen Erfassungssysteme sowie im Bereich der SLAM-Verfahren um eine Lokalisierung und Navigation des mobilen Roboters zu gewährleisten.

DANKSAGUNG

Der Autor möchte sich an dieser Stelle gerne für die fachliche Unterstützung bei Herrn Ing. BSc, MSc Ronald Stärz bedanken. Ferner gilt der Dank Herrn Martin Lederer und der Schlosserei Kreidl GmbH für die Unterstützung bei dem Prototypenbau.

LITERATUR

- [1] SKF. (2022, May) 320/32 x kegelrollenlager. [Online]. Available: <https://www.skf.com/at/products/rolling-bearings/roller-bearings/productid-320%2F32%20X>
- [2] D. J. J. V. Herbert Wittel, Dieter Muhs, *Roloff/Matek Maschinenelemente 20. Auflage*. Springer Fachmedien Wiesbaden GmbH: Vieweg Tubner Verlag, 2011.

- [3] ATO. (2022, May) BLD-50A Controller. [Online]. Available: <https://www.ato.com/Content-doc/BLD50A-DC-Motor-Controller-Manual.pdf>
- [4] EFFEKTA. (2022, May) Batterien btl-reihe. [Online]. Available: <https://www.effekta.com/produkt/btl-reihe/>
- [5] P. Zanuttigh, G. Marin, C. Dal Mutto, F. Dominio, L. Minto, and G. M. Cortelazzo, *Time-of-Flight and Structured Light Depth Cameras*. Cham: Springer International Publishing, 2016.
- [6] B. Heinrich, P. Linke, and M. Glöckler, *Sensorik - Grundlagen Automatisierung: Erfassen - Steuern - Regeln*. Wiesbaden: Springer Fachmedien Wiesbaden GmbH, 2020.
- [7] Pepperl und Fuchs Austria. (2022, september) Ultraschallsensor ub4000-f42-i-v15. [Online]. Available: <https://www.pepperl-fuchs.com/austria/at/classid186.htm?view=productdetailsprodid=24971>
- [8] Intel RealSense. (2022) Intel realsense depth camera 435i. [Online]. Available: <https://www.intelrealsense.com/depth-camera-d435/>
- [9] Robosense. (2022) 3d-lasercanalyzer rs-bpearl. [Online]. Available: <https://www.generationrobots.com/de/403497-3d-lasercanalyzer-rs-bpearl-360-x-90-von-robosense.html>
- [10] Microchip Atmel. (2022, september) Sam v71 xplained ultra evaluation kit. [Online]. Available: <https://www.microchip.com/en-us/development-tool/atsamv71-xult>
- [11] A. Richter, "Verhaltensbasierte koordination eines multi-roboter-szenarios mittels bdi-agenten," Ph.D. dissertation, Bachelor's thesis, Universität Hamburg, Department of Informatics, Germany, 2012.
- [12] A. Klönz, "Softwarearchitekturen der mobilen robotik."
- [13] I. A. Ferguson, "An architecture for dynamic, rational, mobile agents," 1992.



Thomas Lederer ist Student am Masterstudiengang für Mechatronik am MCI in Innsbruck/Österreich. Zudem ist er bei dem Unternehmen NTW Software GmbH als Softwareentwickler im Fachbereich Kommunikationstechnik tätig. Hier insbesondere in der Entwicklung von Aufzeichnungs- und Provisionierungssoftware.

Entwicklung eines Messsystems zur Erfassung der spektralen UV - Intensitäten an einem Quecksilberdampf - Strahler

Dominik Leps, Philipp Augschöll (Betreuer) und Ronald Stärz (Gutachter)

Kurzfassung—Elektroden, welche für den medizinischen Bereich produziert werden, unterlaufen strengsten Qualitätssicherungen, wodurch geringe Abweichungen im Herstellungsprozess zu Ausschuss führen kann. Um die Quote der fehlerhaften Elektroden zu verringern, ist eine umfassende Überwachung des Herstellungsprozesses notwendig.

Für den Herstellungsprozess einiger Elektroden werden ausgehärtete Gele verwendet, welche den Leitwert von der Elektrode zum Körper verringern. Diese Gele werden unter UV - Strahlung zur Polymerisation gebracht. Dabei können Schwankungen der Leistung, in den einzelnen Spektralbereichen der UV - Strahlung, den Prozess stören. Um diese Schwankungen auszugleichen und mögliche andere Umwelteinflüsse auf den Prozess der Polymerisation zu verringern, müssen diese erst erfasst werden. Durch die Verwendung eines Quecksilberdampf - Strahlers als Quelle der UV - Strahlung, wird der Prozess zusätzlich durch den Alterungsprozess des Strahlers beeinflusst. Die vorliegende Arbeit behandelt die Entwicklung eines Messsystems zur Erfassung der spektralen UV - Intensitäten an einem Quecksilberdampf - Strahler.

Die Auswahl und Auslegung der Sensorik basiert auf einer literarischen und firmeninternen Recherche, bezüglich für die Monitoring - sowie Trendaufzeichnungen von Quecksilberdampfstrahlern anwendbaren Messmethoden. Die Analyse des Messumfelds und die damit einhergehenden Messungen von Temperatur und Luftfeuchte, in Kombination mit Simulationen der Elektronik in diesem Messumfeld, bestimmt dabei die Auswahl der Kühlung und Positionierung der Sensorik sowie dessen Auswertelektronik. Der Test des Prototyps erfolgt mit Hilfe einer entwickelten Software zur Auswertung der Sensorik.

Die im Zuge dieser Arbeit entwickelte Auswertesoftware, welche die vom Messsystem erhaltenen Daten bearbeitet, quantifiziert die erhaltenen Messergebnisse,

mit welchen, mittels einer statistischen Auswertung, ein Trend, der den Alterungsprozess der Quecksilberdampf-Strahler beschreibt, erkenntlich gemacht werden kann. Für die Evaluierung einer möglichen Stufenweisen Regelung der Strahler werden jedoch noch Langzeitstudien bezüglich der Polymerisation des Gels benötigt.

Schlagwörter—Polymerisation, UV - Messung, Licht - Messung, EKG-Elektroden

I. EINLEITUNG

ELEKTRODEN werden im medizinischen Bereich sowohl zur Überwachung, als auch zur Behandlung von Patienten verwendet. Ohne dessen Anwendung würden viele medizinische Praktiken nicht funktionieren. Die Hauptaufgabe der Elektroden ist es einen hohen elektrischen Leitwert zwischen dem leitenden Draht und der Haut des Patienten aufzubauen. Um dies zu bewerkstelligen besteht die Seite der Elektrode, welche auf die Haut aufgebracht wird, meist aus Gelen.

Eine Variante, die Gelschicht mit den anderen Elektrodenkomponenten zu vereinen, ist eine durch UV-Strahlung initiierte Aushärtung (Polymerisation). Dieser Prozess ist abhängig von der Übereinstimmung des UV - Strahler - Spektrums und des Absorptions - Spektrums, des im Gel enthaltenen Initiators.

Der Prozess der Polymerisation wird bei Leonhard Lang GmbH, welche Elektroden für den medizinischen Gebrauch herstellen, mithilfe eines UV - Quecksilberdampf - Strahlers, durchgeführt. Dessen Alterungsprozess führt zu Schwankungen im UV - Spektrum, was zu fehlerhaften Elektroden führen

kann. Um diesen Prozess besser zu Überwachen und die Anzahl von fehlerhaften Elektroden zu reduzieren, ist eine automatische Messung der Intensitäten des UV - Spektrums und eine softwaregestützte Erkennung der Alterungserscheinungen des UV - Strahlers ein naheliegender Ansatz.

Die Möglichkeiten das UV - Spektrum eines UV - QD - Strahlers zu erfassen, reichen prinzipiell von Messsystemen, welche nur eine wirkungsbezogene Messung in den Photoinitiatorspektren durchführen können, bis hin zu Messsystemen, welche sowohl das gesamte UV - Spektrum, als auch Messungen im UV - Visible (VIS) - Bereich durch eine Messung abdecken. Folgende im industriellen Bereich erhältliche Messsysteme für UV - Anwendungen sind:

- Das Spektralradiometer UV Spectro von UV - Technik Meyer GmbH. [1]
- Das Handgerät H11 in Kombination mit dem UV - Sensor S11 von UV - Technik Meyer GmbH. [2]
- Das Breit - band Radiometer OmniCure R2000 von Lumen Dynamics Group Inc. [3]

Die Messbereiche der Messgeräte unterscheiden sich von einander, wobei

- das UV Spectro einen Messbereich von 200 nm bis 400 nm Wellenlänge,
- das Handgerät H11 in Kombination den UV - Sensoren S11 einen Messbereich von 230 nm bis 410 nm Wellenlänge und
- das OmniCure mit einem Messbereich von 250 nm bis 1000 nm Wellenlänge hat.

Hierbei bieten alle Messgeräte einen, für die Messung des UV Spektrums des UV - QD - Strahlers, ausreichenden Messbereich. Jedoch muss beachtet werden, dass die Messung des UV Spectros durch einschieben in das Materialband durchgeführt wird, was bei dieser Anlage zu Problemen bei der Produktion führen kann. Zudem kann eine immer gleiche Positionierung auf dem Materialband nicht gewährleistet werden, was zu Messergebnissen führt welche nicht vergleichbar sind.

Des Weiteren muss erwähnt werden, dass um den

gewünschten Bereich der Messung mit dem H11 der Sensor S11 mehrmals getauscht werden muss, wodurch das Selbe Problem wie bei dem UV Spectro Messgerät auftritt. Zudem gibt das H11 nur einen Wert der Bestrahlungsstärke über den kompletten gemessenen Spektralbereich aus, was für eine Trend Erkennung bei den UV - Strahlern unzureichend ist. Zudem ist eine automatische Speicherung der Messdaten nicht möglich

Das OmniCure bietet hingegen einen hinreichendes Messergebnis. Dieses kann jedoch nicht automatisch an die Firmenserver gesendet werden. Zudem müsste für die Messung aller bei der Firma Leonhard Lang befindlichen Anlagen ein Messgerät angeschafft werden, was bei Beschaffungskosten im fünfstelligen Bereich das Budget des Messsystems überschreiten würde. Die Messgeräte oder dessen Aufbauten sind zudem auch nicht für die in der UV - Anlage herrschenden Temperatur von bis zu 98.51° Celsius, welche in Abbildung 1 dargestellt ist, ausgelegt.

II. HARDWARE UND SOFTWARE SETUP

Die Entwicklung des Messsystems lässt sich in drei Abschnitte unterteilen. Dem Hardware Setup, in welchem auf die Auswahl der mechanischen und elektrischen Komponenten, sowie den Aufbau der Platinen des Messsystems eingegangen wird. Den Software Aufbau des Mikrocontrollers in welchen der Ablauf der Software des Mikrocontrollers behandelt wird. Dem Aufbau der Auswertesoftware welche den Ablauf der Auswertung der Messdaten behandelt.

A. Hardware Setup

1) *Mechanischer Aufbau:* Die Mechanischen Komponenten des Messsystems bestehen aus

- einem Lichtwellenleiter,
- einer Halterung der Lichtwellenleiter,
- einem Shutter,
- einem Wellrohr und
- einem Gehäuse.

Der Aufbau der Lichtwellenleiter mit der der Halterung und einem Shutter ist notwendig um die

gegenüber Umwelteinflüssen empfindliche Messelektronik Ausserhalb der UV - Anlage zu montieren. Hierbei ist die Temperatur, mit einer gemessenen Höchsttemperatur von 98.51° Celsius, einer der Umwelteinflüsse, welcher zum Versagen der Elektronik führen kann. Der Verlauf der Temperatur in einer UV - Anlage ist in Abbildung 1 dargestellt.

Der Shutter, welcher an der Halterung der Lichtwellenleiter liegt, wird nur bei durchführung einer Messung geöffnet um die Lebensdauer der Messelektronik zu verlängern. Das Gehäuse und das Wellrohr schützen die Messelektronik sowie die Lichtwellenleiter vor Verschmutzung, elektromagnetischen Einflüssen und physischer Einwirkung.

Die Auswahl dieser Komponenten stützt sich auf den im Lager von der Firma Leonhard Lang GmbH befindlichen Komponenten. Dies kommt zum Einen dem kosteneffizienten Aufbau des Messsystems und zum Anderen einer Austauschbarkeit beschädigter Komponenten zu gute.

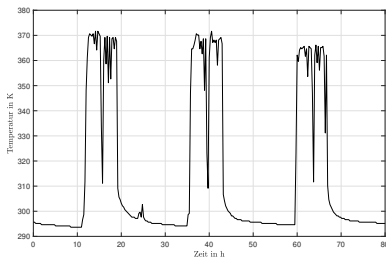


Abbildung 1. Messung der Temperatur in der UV - Anlage [4]

2) *Elektrischer Aufbau:* Um das Messsystem vor Verschmutzung und den im Arbeitsbereich vorkommenden physischen sowie elektromagnetischen Einflüssen zu schützen werden

- ein Harting Stecksystem als Schnittstelle,
- ein Wellrohr als Kabelführung und
- ein Kleingehäuse verwendet.

Um für mögliche zukünftige Erweiterungen des Messsystems einen variablen Aufbau zu gewährleis-

ten werden Klemmen für die Verteilung der Potentiale verwendet. Die Spannungsversorgung des Systems erfolgt über ein in der Produktionsanlage verbautes Netzteil.

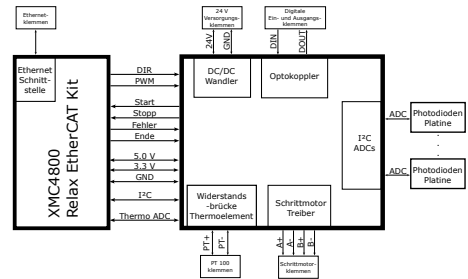


Abbildung 2. Aufbau der Platinen

3) *Platinen Aufbau:* Die Elektronik des Messsystems besteht aus drei verschiedenen Platinen, welche schematisch in Abbildung 2 dargestellt sind. Dem XMC4800 Relax EtherCAT Kit, welches die Steuerungstechnischen Aufgaben übernimmt. Zudem übernimmt es die Aufgabe der Kommunikation mit den Servern der Firma Leonhard Lang über dessen Ethernet Schnittstelle. Das XMC4800 Relax EtherCAT Kit ist mittels Pin Header mit der mittleren Platine verbunden. Auf dieser Platine befinden sich

- die DC / DC Wandler, welche die Versorgungsspannung von 24 V Gleichspannung auf 5 V und 9 V Gleichspannung transformieren,
- die Optokoppler, welche die Digitalen Ein- und Ausgangs Verbindungen mit der SPS der Anlage galvanisch vom restlichen System trennen,
- die Widerstandsbrücke des PT100 Sensors,
- die Ansteuerung des Schrittmotors [5] und
- die Messschaltungsverstärker mit dessen analog zu digital Konverter (ADC), welche den generierten Strom der Photodioden misst. [6]

Um bei Beschädigung einer Photodiode diese möglichst schnell zu ersetzen und um wenn nötig eine Photodiode mit einem anderen Spektralbereich zu verbauen sind die Photodioden auf separaten steckbaren Platinen verlötet.

B. Software Aufbau Mikrocontroller

Die Software des Mikrocontrollers XMC4800 wird mittels der auf IDE DAVE von Infineon Technologies AG umgesetzt. Diese IDE verwendet so genannte DAVE Apps, welche vorgefertigte Programmierfunktionen nutzen um gewisse Funktionen des Mikrocontrollers umzusetzen. Zur Umsetzung dieses Projektes sind folgende Apps in Verwendung:

- DIGITAL_IO wird zur Erfassung und Setzung der Digitalen Ein - und Ausgänge verwendet.
- PWM_CCU4 wird zur Generation eines PWM - Signals verwendet.
- TIMER wird zur Generierung von Wartezeiten zwischen Funktionen verwendet.
- ADC_MEASUREMENT wird zur Analog Digital Konvertierung von an ADC Eingängen anliegende zu messende Spannungen verwendet.
- SPI_MASTER wird zur Generierung der SPI Kommunikation mit den XMC4800 externen ADCs verwendet.
- ETH_LWIP wird zur Generierung der Netzwerkkommunikation via dem UDP Protokoll verwendet.

Der Ablauf des für das Messsystem konzipierte Mikrocontroller Programms ist im Flussdiagramm in Abbildung 3 dargestellt.

Nach dem Start der Software wird auf die Freigabe der Messung, welche von dem digitalen Ausgang der SPS der Produktionsanlage gegeben wird, gewartet. Wenn ein logisches HIGH am digitalen Eingang des Start Signal Pins anliegt wird der nächste Prozess des Programmes ausgeführt. Ansonsten wird die Abfrage des digitalen Eingangs wiederholt.

Der nächste Prozess testet die Netzwerkverbindung des Mikrocontrollers. Um sicher zu gehen das die Kommunikation des Mikrocontrollers mit dem Firmennetzwerk fehlerfrei funktioniert sendet dieser eine Testnachricht an das Netzwerk via dem mit der ETH_LWIP DAVE App ausgeführte UDP Protokoll. Diese Testnachricht wird an den Mikrocontroller zurückgeschickt. Beim erhalten dieser Testnachricht wird diese mit der gesendeten Testnachricht vergli-

chen und somit auf Fehler untersucht. Bei der Erkennung eines Fehlers wird die FehlerUDPKOMvar Variable auf logisch HIGH gesetzt, ansonsten ist diese auf logisch LOW. Der Digitale Ausgang Messsystem_busy wird via der DIGITAL_IO DAVE App auf logisch HIGH gesetzt und der Digitale Ausgang Messsystem_Fehler auf logisch LOW zurück gesetzt.

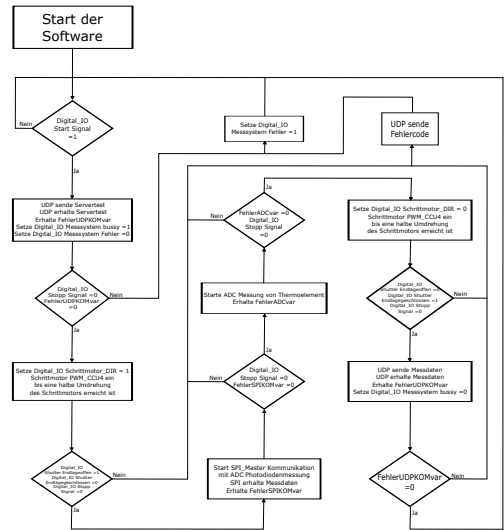


Abbildung 3. Flussdiagramm der Software des Messsystem

Anschließend wird der Prozess auf Fehler überprüft. Bei Auftreten eines Fehlers oder bei Erhalt eines Stopp_Signals wird die Fehlerkette abgehandelt, bei welcher der Digitale Ausgang Messsystem_Fehler wird auf logisch HIGH gesetzt und / oder der Fehlercode auf das Netzwerk gesendet wird. Wenn kein Fehler aufgetreten ist wird der nächste Prozess abgehandelt. Diese Entscheidung wird nach jedem Prozess getroffen.

Der digitale Ausgang Schrittmotor_DIR des Mikrocontrollers wird auf logisch HIGH gesetzt und der PWM Generator, der PWM_CCU4 DAVE App, welcher ein PWM Signal an einem PWM fähigen Ausgang des Mikrocontrollers generiert wird einge-

schalten. Wenn mit den Schritten des PWM Signals eine halbe Umdrehung des Schrittmotors erreicht ist, wird der PWM Ausgang deaktiviert.

Mit Hilfe der SPI_MASTER DAVE App wird eine SPI Kommunikation mit den externen ADCs, welche die Spannung an den Photodioden messen, aufgebaut. Diese werden über einen Befehl zurückgesetzt. Die Kommunikation zwischen dem Mikrocontroller und den ADCs wird mittels Testdaten welche in ein Register der ADCs geschrieben und wieder Ausgelesen werden. Die gesendeten sowie die erhaltenen Testdaten werden miteinander verglichen. Bei der Erkennung eines Fehlers wird die FehlerSPIKOMvar Variable auf logisch HIGH gesetzt, ansonsten ist diese auf logisch LOW. Anschließend wird die ADC Messung via SPI gestartet. Der Speicher der ADCs wird mit einem weiteren Befehl ausgelesen. Der Parity check der Messdaten, welcher anschließend ausgeführt wird, setzt bei einem Fehler die FehlerSPIKOMvar Variable auch auf logisch HIGH.

Im nächsten Prozess werden die Messungen am Thermoelement von den zwei ADC Kanälen, welche mit der ADC_MEASUREMENT DAVE App konfiguriert sind, durchgeführt. Bei einem Fehler in dieser Messung wird die FehlerADCvar Variable auf logisch HIGH gesetzt, ansonsten ist diese auf logisch LOW.

Anschließend wird der Shutter des Messsystems wieder geschlossen, was mit der Setzung des digitalen Ausgang Schrittmotor_DIR des Mikrocontrollers auf logisch LOW und einer Einschaltung des PWM Ausgangs realisiert ist. Wenn mit den Schritten des PWM Signals eine halbe Umdrehung des Schrittmotors erreicht ist, wird der PWM Ausgang deaktiviert.

Im Letzten Prozess wird digitale Ausgang Messsystem_bussy auf logisch LOW gesetzt und die gespeicherten Messdaten an das Firmeninternen Server gesendet. Diese senden die erhaltenen Daten an den Mikrocontroller zurück. Dieser vergleicht die erhaltenen Daten mit den gesendeten und setzt je nach Ergebnis die FehlerUDPKOMvar Variable auf logisch HIGH oder LOW. Bei Erhaltung eines Fehlers wird der dementsprechende Fehlercode an das Netzwerk

gesendet, der Digitale Ausgang Messsystem_Fehler wird auf logisch HIGH gesetzt und es wird auf ein Start Signal der SPS gewartet. Wenn die Sendung der Daten fehlerfrei abgelaufen ist wird kein Fehler gesendet und direkt auf das Warten auf ein Startsignal der SPS übergegangen.

C. Aufbau Auswertesoftware

Die Software ist mit der IDE MATLAB der Firma MathWorks Inc. realisiert. Der Schematische Aufbau der Software ist im Flussdiagramm in Abbildung 4 dargestellt. Das Ziel dieser Software ist es, die erhaltenen Messdaten, welche durch die Anwendung von Lichtwellenleiter gedämpft sind, so zu modifizieren, dass diese einer Messung in der UV - Anlage gleichen. Eine weitere Funktion dieser Software ist es die erhaltenen und Modifizierten Daten der Messung zu Quantifizieren um mögliche Trends in der UV - Strahlung der UV - Anlage zu erkennen, wie die Verunreinigung der Strahler. Die Hauptfunktion der Auswertungssoftware ist es, durch Quantifizierung dieser Messergebnisse einen Trend zu erkennen welcher die Alterungserscheinung der UV - QD - Strahler aufzeigt.

Nach dem Start der Auswertesoftware wird diese im ersten Prozess initiiert. Hierbei wird aus den Datenblättern der Photodioden das Verhalten der Sensoren in deren Spektralen Messbereich, aus dem Graphen, welche die vom Hersteller angegebenen Empfindlichkeits Kurven enthalten, mit Hilfe eines dafür ausgelegten Bildverarbeitungs Programms, entnommen. Zur Extraktion der Ausgangsmessung aus den nur im Bildformat vorliegenden Messdaten der UV - Intensitäten über das gesamte UV - Spektrum, welches in den verschiedenen UV - Anlagen variiert, ist ein weiteres dafür ausgelegtes Bildverarbeitungs Programm in Verwendung. Die in diesem Prozess erfassten Sensorfunktionen und die Ausgangsmessung bilden die Basis für die weiteren Berechnungen der Auswertesoftware.

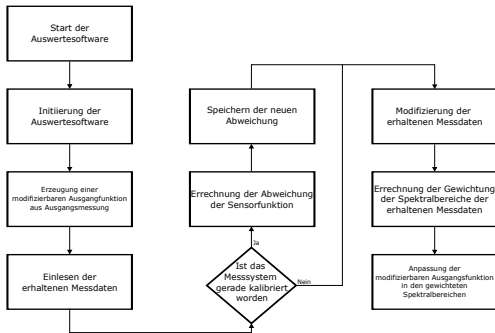


Abbildung 4. Flussdiagramm der Software zur Auswertung der Messdaten

Im nächsten Prozess der Auswertesoftware, wird aus der Ausgangsmessung, welche mit einem externen Messgerät aufgenommen ist, durch die Anwendung mehrerer Funktionen, eine modifizierbare Ausgangsfunktion erzeugt. Diese Funktion ist über deren markantesten lokalen Maxima modifizierbar. Die Ausgangsfunktion ist über die Höhe, die Breite und die Erhebung über der Basis der lokalen Maxima steuerbar.

Nach dem einlesen der erhaltenen Messdaten ist die letzte Messung mittels einem externen Messgerätes, welches zur Kalibrierung der Auswertesoftware in Verwendung ist, relevant für den weiteren Verlauf der Auswertesoftware. Wenn diese Messung zur selben Zeit wie die der erhaltenen Messdaten getätigt wird, wird die Abweichung der Sensorfunktion mit der neuen Ausgangsmessung neu berechnet und für weitere Modifizierungen der erhaltenen Messdaten gespeichert. Andernfalls werden die erhaltenen Messdaten direkt mit der bereits gespeicherten Abweichung modifiziert.

Im folgenden Prozess werden die Gewichtungen der aufgeteilten Spektralbereiche der Sensoren berechnet, welche sich aus der Kombination der Ausgangsmessung und der Sensorfunktionen mit den selben Aufteilungen der Spektralbereiche ergeben.

Im letzten Prozess wird die Ausgangsfunktion in den Spektralbereichen mittels eines Algorithmus mo-

difiziert. Die Suchfunktion des Algorithmus basiert auf der Fehlerminimierung in den einzelnen Spektralbereichen. Dieser Fehler entsteht aus der Differenz des theoretischen Messergebnisses im Bezug auf die gemessenen Daten multipliziert mit der theoretischen Gewichtung. Das theoretische Messergebnis ergibt sich dabei aus der modifizierten Ausgangsfunktion.

Bei Erhalt eines neuen Messergebnisses werden die Prozesse, welche im Flussdiagramm in Abbildung 4 dargestellt sind, erneut durchlaufen. In den folgende Unterkapiteln sind diese Prozesse detaillierter behandelt.

III. ERGEBNISSE

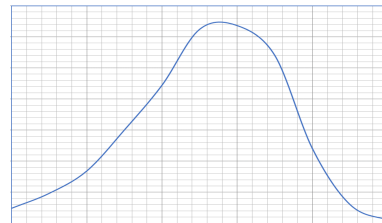


Abbildung 5. Bildausschnitt der Empfindlichkeits Kurve der Photodiode der Type SD008-2171-112 aus [7] entnommen.

In Abbildung 5 ist die Empfindlichkeits Kurve der SD008-217-112 Photodiode der Firma Advanced Photonix Inc. dargestellt. Diese Abbildung wurde aus dem Datenblatt der Advanced Photonix SD008-217-112 Photodiode entnommen und ist die einzige Quelle der Empfindlichkeitskurve dieses Sensors. Die händische Entnahme der Empfindlichkeitskurve aus dieser Abbildung ist mit einem unverhältnismäßig hohen Aufwand verbunden. Insbesondere wenn dies für mehrere verschiedenen Sensoren gemacht werden muss. Wie in dieser Abbildung erkenntlich, sind zur einfachen Extraktion der Werte, via auslesen der Pixel der Kurve, zu viele Störende Elemente wie das Raster und der Rahmen enthalten.

Mit der in Aufbau der Auswertesoftware beschriebenen Funktion können die störenden Elemente entfernt und die Daten der Empfindlichkeits Kurve der

SD008-217-112 Photodiode aus dem Bild in Abbildung 5 entnommen werden. Zudem wird ebenfalls eine mit der Matlab Funktion *polyfit* an diese Daten angepasste Polynomfunktion ausgegeben. Den Vergleich der direkt entnommenen Daten und der Polynomfunktion der Empfindlichkeits Kurve der SD008-217-112 Photodiode ist in Abbildung 6 dargestellt.

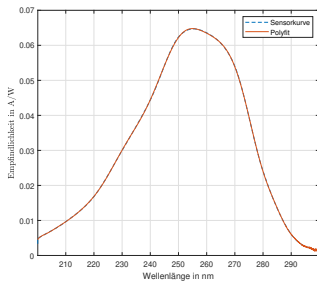


Abbildung 6. Vergleich der Empfindlichkeits Kurve mit der daraus generierten Polynomfunktion

Wie in Abbildung 6 ersichtlich ist der Verlauf der *Sensor Kurve* und die an diesen Daten angepasste Polynomfunktion *Polyfit* fast deckungsgleich, was für eine ausreichende Anpassung der Polynomfunktion an die Werte im Datenblatt spricht. Ab einer Wellenlänge von 290 nm kann man jedoch ein eindeutiges Rauschen bei der Polynomfunktion feststellen. Bei der Berechnung des Auswertungsprogramms kann dieses Rauschen schon zu Problemen bei der Quantifizierung der Messergebnisse führen, was das weitere Verwenden von Polynomfunktionen zur Beschreibung der Empfindlichkeits Kurve des Sensors hinfällig macht.

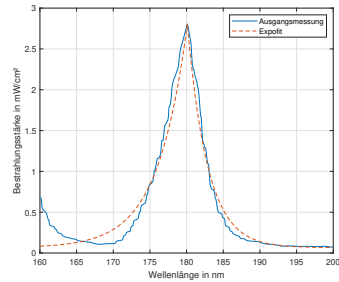


Abbildung 7. Vergleich der *Ausgangsmessung* mit der Exponentialfunktionen *Expofit*

In Abbildung 7 ist der Vergleich der *Ausgangsmessung* mit den an beiden Seiten des lokalen Maximums angepasste Exponentialfunktionen *Expofit* dargestellt. Es ist festzustellen, dass die Anpassung der Exponentialfunktionen im obersten Bereich des lokalen Maximums genauer ist als im unteren, dies hat jedoch keine Auswirkungen auf die Auswertesoftware, da mit diesen Funktionen nur die Änderung der Ausgangsmessung durchgeführt werden.

IV. CONCLUSIO

Die Festlegung des Messsystem-Konzepts auf eine Variante mit Lichtwellenleiter hat den Aufwand des Aufbaus des Messsystem und auch dessen Kosten um ein vielfaches reduziert, da eine aufwändige Temperaturregelung des Messsystems mit Kühlapparaten hinfällig wurde. Des Weiteren wird durch den Aufbau dieser Variante ein einfacher Zugang mit externen Messgeräten, welche zur Kalibrierung des Messsystem verwendet werden können, gewährleistet.

Die Erstellung einer Software zur Extrahierung der Empfindlichkeits Kurve aus einem Datenblatt eines Photodiodenherstellers war aufgrund des fehlenden Standarts der Abbildung dieser Kurven mit sehr Hohen Aufwand verbunden. Durch die geringe Auflösung der Abbildung dieser Kurven waren einige Abbildungen vollkommen unbrauchbar. Die Alternative zu der Erstellung dieser Software war die

Kontaktierung der Hersteller, welche die Messdaten meist nicht zur Verfügung hatten oder die ungenaue langwierige Messung der Kurve per Hand. Bei Fertigstellung dieser Software waren ihre Funktionen für die Extraktion der Meisten Empfindlichkeits Kurven ausreichend. Die an diesen erhaltenen Daten angepasste Polynomfunktion war jedoch für die Verwendung in der Auswertesoftware unbrauchbar.

Aufgrund der Probleme mit den Lieferketten von Elektronikkomponenten war die Auswahl von für dieses Projekt passender Photodioden stark begrenzt, da einige Photodioden nicht mehr auf Lager waren oder nur in großen Stück-mengen erhältlich waren, war eine Zusammenstellung mehrerer Photodioden für die Messkette nicht möglich. Dies hat nicht nur die Fertigstellung der Platine verhindert, sondern auch die Fertigstellung der Auswertesoftware erschwert, da Testmessungen mit dem Messsystem keine Option waren.

Das mittelfristige Ziel ist die Fertigstellung des Prototyps, welcher durch Messungen am UV - QD - Strahler mit Hilfe der Auswertesoftware den Trend des Alterungsprozesses der UV - QD - Strahler erkennen kann. Dies kann umgesetzt werden wenn die gewünschten Photodioden wieder auf Lager sind und in geringeren Stückmengen zur Verfügung stehen. Die großen Stückmengen sind in der Hinsicht das größere Problem, da die Beschaffung einer Photodiодentype, bei einer Mindeststückzahl von 1000 Stk, das Budget der Entwicklung dieses Messsystems sprengen würde. Insbesondere da für den Prototypen zur Testung der Potodioden nur ein bis zwei Stück jeder Photodiодentype benötigt wird und erst nach Implementierung der Messdaten in der Auswertungssoftware ermittelt werden kann, ob diese Photodioden Type für diese Messkette passend ist.

Das langfristige Ziel ist die Vollständige Überwachung des Polymerisationsprozesses in der UV - QD - Strahler - Anlage. Hierfür muss mit Hilfe des konzipierten Messsystems zusätzlich eine Messung der

Luftfeuchte und eine akkurate Messung der Temperatur an mehreren Stellen innerhalb und außerhalb der Anlage durchführbar werden. Dafür wird ein neuer Aufbau der Platinen benötigt. Des Weiteren kann auf der Basis des Messsystems in der Kombination mit der Auswertesoftware, eine Software zur Regelung der UV - QD - Strahler entwickelt werden, was den Prozess der Polymerisation noch stabiler machen würde.

LITERATUR

- [1] uv-technik meyer GmbH, "UV Spectro - D Anleitung," Ortenberg, 2017.
- [2] —, "SI 1 HI 1 - E Data sheet," Ortenberg, 2018.
- [3] Lumen Dynamics Group Inc., "R2000 Radiometer UV/Visible Radiometer User's Guide," USA und Kanada, 2018.
- [4] Leonhard Lang GmbH, "Auswertung Luftfeuchte Temperatur," Österreich, 2021.
- [5] Texas Instruments Inc., "DRV8825 Stepper Motor Controller IC Datasheet," USA, 2020.
- [6] D.-K. Electronics. (2019) Implementierung eines hochempfindlichen spektrophotometrischen messkreises. USA. [Online]. Available: <https://www.digikey.at/de/articles/how-to-implement-a-high-sensitivity-spectrophotometric-sensing-circuit> (accessed: 20.06.2022)
- [7] A. Inc., "Advanced photonix sd008-2171-112 datasheet," USA, 2022.



Dominik Leps ist Student am Masterstudiengang für Mechatronik am MCI Innsbruck/Österreich. Zudem ist er bei Leonhard Lang GmbH als Steuerungstechniker in der Abteilung Maschinenbau und Technik tätig.

Photogrammetrie des Olympia Eiskanals Innsbruck und Implementierung in ein virtuelles Erlebnis

Dominic Ludwig, Bernhard Hollaus (Betreuer)

Kurzfassung—Der Anwendungsbereich und das Interesse an der virtuellen Realität ist in den letzten Jahren stark gestiegen, was nicht zuletzt durch die stetige Weiterentwicklung entsprechender Technologien ermöglicht wurde. Immersive VR-Brillen und Drohnentechnologien sowie leistungsstarke Grafikprozessoren und immer besser werdende Software bieten neue Möglichkeiten, virtuelle Welten zu entwickeln. Moderne Photogrammetriesoftware liefert dabei die nötigen Werkzeuge, um bestehende Objekte und Umgebungen zu virtualisieren.

Der Olympia Eiskanal in Innsbruck-Igls bietet die Möglichkeit, mithilfe der Photogrammetrie eine virtuelle Umgebung nach einem realen Vorbild zu erstellen. Dies soll ferner eine virtuelle Rodelfahrt ermöglichen und kann für verschiedene Zwecke genutzt werden. So wird im Rahmen dieser Masterarbeit ein Workflow entwickelt und umgesetzt, um eine virtuelle Version des Eiskanals und der Umgebung zu erzeugen und zudem eine Rodelfahrt zu simulieren.

Diese Arbeit beinhaltet das Gamification-Konzept für den virtuellen Eiskanal und die Erstellung und Umsetzung des Workflows. Dabei werden mithilfe einer Drohne und weiteren Kameras Bilder aufgenommen und Datensätze erstellt. Die Berechnung der 3D-Modelle findet in Reality Capture statt. Für die Photogrammetrie werden ausschließlich Bilddaten verwendet. Im Postprocessing werden die Modelle dann ausgebessert, da es nicht zu erwarten ist, dass große und komplexe Modelle fehlerfrei berechnet werden. Außerdem werden die 3D-Modelle für die Implementierung und Gamification in Unity optimiert, sodass eine Anwendung auf einer Plattform im mittleren Performancebereich möglich ist. In Unity findet anschließend eine Gamification statt, bei der das photogrammetrische Modell in eine Open-World-Umgebung implementiert wird und interaktive

Komponenten programmiert werden.

Das Ergebnis dieser Arbeit ist ein 3D-Modell des Eiskanals, bestehend aus 8 Teilmodellen. Diese wurden aus etwa 28600 Bildern rekonstruiert. Durch die Gamification wurde eine virtuelle Rodelfahrt umgesetzt und eine ausführbare Anwendung für eine Windows-Desktop-Plattform kompiliert. Die produzierte Datenmenge der Source-Bilder, Datensätze und Projektdaten beträgt ca. 2 TB. Die Ergebnisse und Daten dieser Arbeit bieten die Möglichkeit, dieses Projekt weiter zu optimieren oder ähnlich Projekte auszuführen und letztlich für VR-Anwendungen auszulegen.

Schlagwörter—Photogrammetrie, Unity, Blender, Drohne, Digitalisierung, Gamification, Eiskanal, Bob, Rodel, Skeleton, virtuell

I. EINLEITUNG

A. Motivation

DER Olympia Eiskanal in Innsbruck ist eine kombinierte Kunsteisbahn für Bob-, Rodel- und Skeletonrennen. Eine solche Kombination war zur Zeit der Erbauung im Jahre 1974 eine Neuheit und wirkte als Vorbild für zukünftige Bahnen. Der Kanal wurde anlässlich der Olympischen Winterspiele 1976 gebaut und befindet sich in Innsbruck-Igls. Mit seiner Gesamtlänge von 1270 m und 14 Kurven dient der Eiskanal bis heute für diverse Veranstaltungen und Weltcuprennen. [1]

Den Innsbrucker Eiskanal im Rahmen dieser Arbeit in eine virtuelle Realität zu implementieren bietet die Möglichkeit, moderne Technologien im Software-

und Hardwarebereich zu nutzen. Solche Technologien wie Drohnenaufnahmen und Photogrammetrieprogramme wurden in den letzten Jahren stetig weiterentwickelt und verbessert. Diese ermöglichen es, eine realitätsgetreue Umgebung nach einem bestehenden Vorbild zu erschaffen, in der im Weiteren eine Interaktion mit der Umwelt stattfinden kann. Auf diese Weise können nicht nur Sportarten wie in diesem Fall eine Rodelfahrt simuliert werden. Es können außerdem realistische Umgebungen erzeugt werden, die zusammen mit Multibodymodellen für Echtzeitsimulationen zum Einsatz kommen [2].

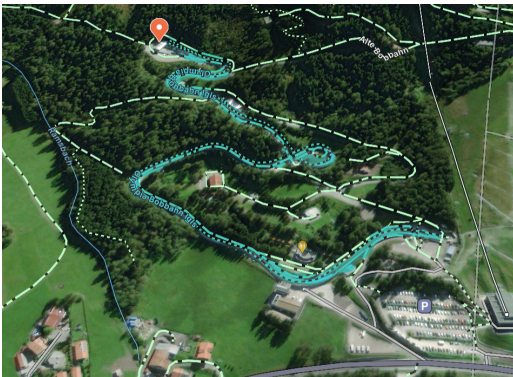


Abbildung 1: Verlauf Eiskanal [3]

Es wird im Rahmen dieser Arbeit ein Workflow entwickelt, um eine solche virtuelle Welt nach einem realen Vorbild zu erzeugen. Der virtuelle Eiskanal soll ferner die Möglichkeit bieten, den Rodelsport besser vermitteln zu können und mehr Interesse zu wecken, indem ein virtuelles Erlebnis geschaffen wird. Das Ergebnis kann beispielsweise in Schulen, bei Veranstaltungen und auf Messen gezeigt werden, mit der Absicht, mehr Menschen für diesen Sport zu begeistern.

B. Vorausgehende Arbeiten

Im Bereich der Spieleentwicklung und Erstellung von virtuellen Welten gab es in den letzten Jahren

einige interessante Arbeiten, welche die Verwendung von Photogrammetrie zur Rekonstruktion dreidimensionaler Objekte und Umgebungen heranziehen.

In der akademischen Fachzeitschrift, *Games and Culture*, erschien 2020 ein Artikel mit dem Titel *Use of Photogrammetry in Video Games: A Historical Overview* [4]. In dem Artikel wird beschrieben, wie die Popularität von Photogrammetrie für Spieleentwickler seit 2014 stark zunimmt. Die Verwendung von Photogrammetrie in dem Spiel *The Vanishing of Ethan Carter*, welches mit detailreichen und realistischen Umgebungen überzeugt, habe in der Industrie für starke Diskussionen gesorgt. Spieleentwickler würden nun, wenn auch später als andere industrielle Bereiche, vermehrt Techniken der Photogrammetrie verwenden und anpassen, um eine nahtlose Einbindung der Objekte in die Spiel-Engines zu ermöglichen.

C. Zielsetzung

Das übergeordnete Ziel dieser Arbeit ist die Erstellung der virtuellen Welt, in der sich der Anwender auf einen Rodel setzen und den Eiskanal herunterfahren kann. Der genaue Ablauf der virtuellen Erfahrung wurde im Vorhinein nicht festgelegt und wird im Rahmen der Arbeit definiert. Das digitale Modell des Eiskanals und der Umgebung wird über eine Photogrammetrie erstellt. Dabei werden ausschließlich Bilder von Luft- und Bodenaufnahmen verwendet. Die interaktive Komponente soll über eine Gamification mit der Spiel-Engine Unity erzielt werden, wobei eine Rodelfahrt durch den Eiskanal simuliert wird.

II. METHODEN

Das Gesamtprojekt gliedert sich grundlegend in fünf Abschnitte (siehe Abbildung 2). Der erste Teil stellt die Einarbeitungsphase dar und beinhaltet Recherchen bezüglich Photogrammetrie und Spieleentwicklung in Unity, sowie die Definition von Randbedingungen für das Projekt in Abstimmung mit der Betreuung. Teil zwei behandelt die Konzeptfindung. Dort werden, unter Berücksichtigung der

Randbedingungen, mögliche Gamification-Konzepte und Vorgehen für die Erstellung der virtuellen Welt angeführt und verglichen. Mit dem Ergebnis der Konzeptphase kann die Umsetzung begonnen werden. Teil drei beschreibt die Umsetzung der Photogrammetrie. Hier ist die Erstellung der virtuellen Umgebung dokumentiert. Teil vier dokumentiert den Prozess der Gamification, in dem die virtuelle Welt mit einer interaktiven Komponente verbunden wird. Im fünften und letzten Teil wird der Projektabschluss sichergestellt und die Dokumentation fertiggestellt.

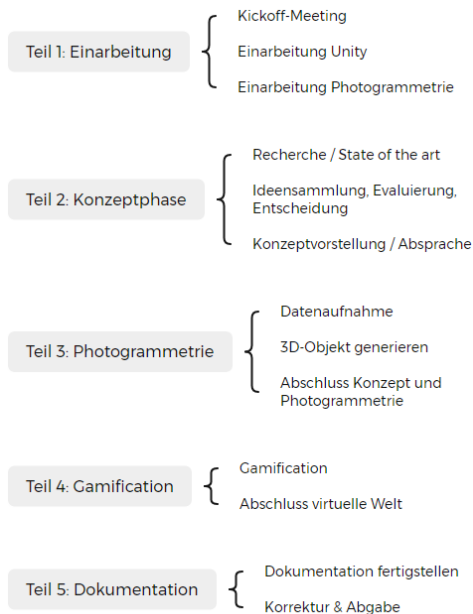


Abbildung 2: Projektübersicht

A. Konzept

Unity Game-Engine [5] bietet mit ihrer Plattform eine Laufzeit und Entwicklungsumgebung für Echtzeitanwendungen. Mit ihrer leistungsstarken Spiel-Engine, ausführlichen Dokumentation [6] und einer großen Community ist sie eine der bekanntesten und

etabliertesten Entwicklungspalttformen für Spiele-Entwickler. Weitere bekannte Spiel-Engines sind Unrealengine [7] und Cryengine [8]. Unity Skripte basieren auf C# als Programmiersprache und Spiele können für verschiedene Plattformen kompiliert werden. Mögliche und in diesem Kontext relevante Spiel-Konzepte, welche Unity standardmäßig anbietet sind: (2D), 3D, 2D/3D Mobile, AR, VR, Third Person (3D), First Person (3D). Weitere Funktionen und Module können in der Dokumentation [6] nachgelesen werden.

Für dieses Projekt wird ein 3D-Projekt in Unity's Universal Render Pipeline (URP) [9] erstellt. URP basiert auf der Scriptable Render Pipeline (SRP) und kann für ein großes Spektrum an Plattformen eingesetzt werden. So können zum Beispiel auch Anwendungen für Plattformen mit niedriger Performance, wie VR-Headsets erstellt werden. Diese können bei der Verwendung der High Definition Render Pipeline (HDRP) schneller in Performanceprobleme laufen, da die Anwendung doppelt (einmal je Auge) gerendert werden muss und im Fall einer Standalone-Plattform weniger Grafikperformance vorhanden ist. Da die VR-Anwendung in diesem Fall sehr interessant ist und die URP dahingehend sehr flexibel ist, wird diese gewählt.

B. Photogrammetrie

Die Photogrammetrie bietet die nötigen Mittel, um mithilfe von 2D-Bildern dreidimensionale Modelle zu erstellen. In diesem Fall handelt es sich um eine Nahbereichs-Photogrammetrie, bei der die Bilder aus einer Entfernung von maximal 50 Meter gemacht werden. Dazu werden Luftaufnahmen von einer Drohne und Detailaufnahmen von einer Handkamera verwendet. Die eigentlichen Berechnungsschritte finden dann in einer Software (Reality Capture) statt. Im Rahmen dieser Arbeit wird ein Gelände von ca. 5,7 ha erfasst, welches den Eiskanal an sich und mehrere Gebäude beinhaltet. Zu sehen ist das Gelände in Abbildung 3.



Abbildung 3: Areal Unterteilung [10]

Tabelle I: Teilmodelle

Abschnitt	Inhalt
1	Starthaus
2	Kurve 1
3	Kurve 2
4	Kurve 3, 4, 5
5	Kurve 6
6	Kurve 7, 8, 9
7	Kurve 10, 11, 12
8	Kurve 13, 14, Ziel/Auslauf

Wie in der Abbildung zu erkennen ist, wird das Gesamtmodell in insgesamt 8 Teilmodelle aufgeteilt. Für die Aufteilung gibt es mehrere Gründe:

- 1) Durch die Unterteilung in kleinere Teilmodelle können die nötigen Fotos meistens innerhalb eines Tages gemacht werden.
- 2) Die Berechnungsdauer für Alignment, Modellberechnung und Texturierung steigt mit der Anzahl der Bilder an und es werden Hardware-Limitierungen erreicht.
- 3) Kleinere Teilmodelle führen zu einem höheren Detailgrad und einem qualitativ besseren Modell, sowohl im Hinblick auf die Modellerstellung als auch für die Verwendung in Unity. So ermöglichen mehrere kleine Gameobjects in Unity die Anwendung von LOD [11] und Occlusionculling [12]. Außerdem ist sowohl die Darstellung des 3D-Modells in RC, als auch das Arbeiten mit diesem in 3D-Software wie Blender stark abhängig von den Hardwarekomponenten – vor allem GPU und RAM – und somit limitiert durch die Anzahl der Flächen im Modell.

In folgender Tabelle sind die Abschnitte zusammengefasst:

1) *Theorie*: Die digitale Photogrammetrie kann im Grunde eine Reihe an verschiedenen Daten verwenden. Dabei können sowohl aktive (z. B. Laser-Scanning, Radar) als auch passive (z. B. Fotografie) Sensoren zum Einsatz kommen [13]. In dieser Arbeit werden ausschließlich Daten aus Bildsensoren von Fotokameras verwendet. Wie bereits angesprochen, wird aus einem Satz von Bildern ein dreidimensionales Objekt erstellt. Dazu werden zunächst die einzelnen Kamerapositionen im Raum bestimmt. Dieser Prozess wird in den meisten Programmen als *Camera-Alignment* bezeichnet. Der überlappende Teil (Overlap) eines Bildpaares, auch Stereopaar genannt, liefert dabei die nötigen stereoskopischen Parallaxe [14, S. 57 ff.], die für die Berechnung verwendet werden [13].

In der Software muss also eine *Feature-Detection* auf Pixelebene stattfinden, um Parallelen zwischen verschiedenen Bildern, genauer gesagt gleiche Features innerhalb eines Stereopaars zu finden und so ein Bildpaar zu matchen [15, S. 40]. Abbildung 4 zeigt die Visualisierung des Feature-Matchings in Reality Capture. Zu sehen sind drei verschiedene Bilder aus leicht unterschiedlichen Blickwinkeln mit einem großen Overlap. In dem unteren linken Bild sind die Features (in RC: Tie-Points) als Grüne Punkte abgebildet. Die Linien führen zu denselben Features in den beiden anderen Bildern.

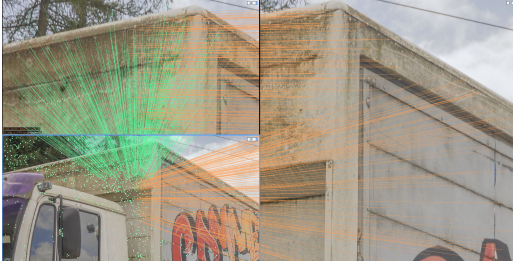


Abbildung 4: Feature detection und Tie-Points in Reality Capture

Stehen die Kamerapositionen im Raum fest, wird eine Punktwolke berechnet. Dabei können ausschließlich Punkte bestimmt werden, die auf mindestens zwei Bildern zu sehen sind. Aus der Punktwolke wird anschließend die Oberfläche des Modells (Mesh) berechnet, indem die einzelnen Punkte zu Flächen – im Normalfall sind das Dreiecke – und so zu einem Netz miteinander verbunden werden.

2) *Equipment:* Die Luftaufnahmen werden mit einer Drohne vom Typ DJI Mavic 3 gemacht. Die Kamera ist an einem 3-Achs-Gimbal befestigt. Es handelt sich um eine Hasselblad-Kamera mit 3/4 CMOS Sensor und 20 Megapixel. An die Kontrolleinheit wird ein Smartphone für die Steuerung und Videoübertragung angeschlossen.



Abbildung 5: Equipment, Drohne [16]

Des Weiteren wird für die Bodenaufnahmen eine digitale Spiegelreflex-Kamera (DSLR) vom Typ Sony SLT A58 verwendet. Diese Kamera hat ebenfalls eine

Auflösung von ca. 20 Megapixel und nutzt einen etwas größeren CMOS-Sensor im APS-C Format. Die gesamte Hardware für die Photogrammetrie ist in folgender Tabelle dargestellt:

Tabelle II: Hardware & Software

Bezeichnung	Modell
Kameras & Drohne	Sony SLT A58, Gopro 9, DJI Mavic 3
CPU	Intel Core i7 10700K
GPU	Nvidia GTX 1050ti
VRAM	4 GB
RAM	64 GB DDR4
Projekt-Speicher (aktive Projekte)	1 TB NVMe SSD-Speicher
Archiv-Speicher (fertige Projekte, Source-Bilder)	4 TB SSD-Speicher
Backup-Speicher	2 TB Cloud-Speicher
Photogrammetriesoftware	Reality Capture (RC)
Bildbearbeitung	Adobe Camera Raw, Adobe Photoshop
3D-Modellierung	Blender

3) *Workflow:* Der Workflow für die Erstellung der jeweiligen Teilmodelle ist in Abbildung 6 zu sehen. Im Folgenden werden diese Schritte näher erläutert.

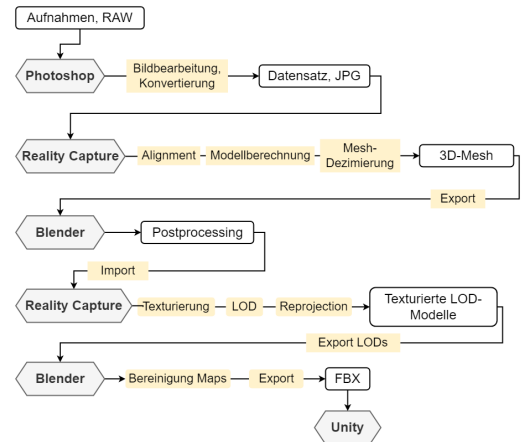


Abbildung 6: Photogrammetrie Workflow

a) *Aufnahmen:* Als Erstes werden die Aufnahmen gemacht. Diese werden im RAW-Format aufgenommen, um bessere Einstellmöglichkeiten für die Bildbearbeitung zu ermöglichen. Das Ziel ist es, möglichst viele Bilder mit einem großen Overlap zu produzieren. Dabei sollte darauf geachtet werden, einen möglichst durchgängigen Verlauf der Kamera zu haben. Das erleichtert dem Programm die Berechnung der Kamerapositionen im Vergleich zu einem Datensatz, dessen Stereopaare getrennt über den gesamten Datensatz verteilt sind.

b) *Datensatz:* Die Erstellung des Datensatzes ist ein wichtiger Arbeitsschritt in diesem Prozess. Ein guter Datensatz liefert ein unkompliziertes und schnelles Alignment und ein gutes Endresultat in Bezug auf die Netz- und Texturqualität.

Zunächst werden unscharfe, verschwommene und doppelte Bilder sowie Bilder mit unerwünschten Objekten oder durchlaufenden Menschen aussortiert. Anschließend werden die Bilder über die Batch-Processing-Funktion in Adobe Camera Raw mit einer passenden Voreinstellung versehen und in das JPG-Format konvertiert. Hier können Kontrast verringert sowie Höhen und Tiefen ausgeglichen werden, um Schatten und Reflexionen zu minimieren.

c) *Kamera-Alignment:* Für das Alignment wird ein Komponentenworkflow [17] verwendet, bei dem das Alignment zunächst für Teilkomponenten stattfindet. Diese Teilkomponenten können als *.rcalign* Datei exportiert und zusammen in ein neues Projekt importiert werden. Dort findet dann das Alignment für das gesamte Modell statt. Diese Methode kann vor allem für große Projekte Vorteile bieten, wie Zeiterparnisse im Alignment-Prozess und ein schnelleres Ergebnis für Datensätze, welche die Software nicht als Ganzes in eine Komponente ausrichten kann und dadurch in Einzelkomponenten aufteilt.

d) *Berechnung:* Die Berechnung bezieht sich hier auf die Rekonstruktion des 3D-Modells. Dabei werden in RC zwei Phasen durchlaufen, die Berechnung der Depthmaps und die Erstellung des Netzes. Berechnungszeiten sind in Tabelle III zusammengefasst.

e) *Postprocessing:* Das Postprocessing besteht im Grunde darin, Fehler im Mesh auszubessern und das Modell auf eine Größe zu reduzieren, mit der ein durchschnittlicher PC arbeiten kann.

f) *Texturierung:* Um dem Modell Farbe und Struktur zu verleihen, muss eine Textur erstellt werden, diese wird ebenfalls in RC berechnet. Dazu generiert das Programm eine UV-Map [18] – eine ebene Repräsentation des abgewickelten 3D-Meshs – und projiziert die Bilder auf diese Map. So werden ferner die Farben der Bilder auf das Mesh projiziert. Diese Map wird auch als Base-Map bezeichnet.

Neben den Farben der Bilder können außerdem weitere Informationen auf die UV-Map projiziert werden, wie zum Beispiel Informationen über Flächenrichtungen und Verschiebungen. Diese werden dann als Normal- und Height-Map bezeichnet [19]. Normal- und Height-Maps dienen dazu, geometrische Details von den Ursprungsmodellen auf stark reduzierte Modelle zu übertragen. In dieser Arbeit werden Normal-Maps lediglich für das oberste LOD-Modell (LOD 0) verwendet.

g) *Reprojection:* RC kann die berechneten Texturen von einem Modell auf das andere übertragen (Reprojection). Diese Funktion wird verwendet, um einerseits die Textur auf die LOD-Modelle zu übertragen und dabei Zeit zu sparen, andererseits wird diese Funktion genutzt, um eine Normal-Map für das oberste LOD-Modell zu erstellen.

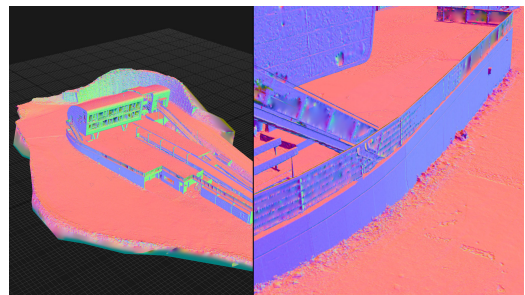


Abbildung 7: Darstellung Normal-Map in RC

Dabei werden die Informationen der Flächennor-

malen des Ursprungsmodells auf das vereinfachte Modell übertragen, um Oberflächenbeschaffenheiten und verloren gegangene Details der Geometrie darstellen zu können. Da die Normal-Map aus dem ursprünglich berechneten Modell generiert werden muss, um dessen Detail zu übernehmen, werden auch Fehler im Mesh mit übertragen. Diese müssen in einer Bildbearbeitungssoftware ausgebessert werden (siehe II-B3h, Punkt 4). In Abbildung 7 ist der Normal-Layer in RC dargestellt. Hier ist zu erkennen, dass sowohl die Details wie die Struktur von Mauer, Boden und Zaun dargestellt werden können, aber auch die Fehler, wie Löcher im Mesh des Ursprungsmodells übertragen werden.

h) *Finaler Export*: Ein Teilmodell mit allen drei LOD-Meshes und deren Texturen soll jeweils in einer einzigen Datei gespeichert werden. Dazu wird das *.fbx* Format gewählt. Damit Unity die LOD-Layer erkennt, müssen die Modelle korrekt exportiert werden. Außerdem sollen die Teilmodelle bereits zueinander positioniert werden, sodass eine Positionierung in Unity nach dem Import nicht mehr nötig ist. Dies wurde folgendermaßen gelöst:

- 1) Alle drei LOD-Modelle werden als *.obj* Datei, inklusive deren Texturen exportiert und in ein neues Blender-Projekt importiert.
- 2) Hier werden alle Meshes auf unterschiedliche Layer gelegt und wie folgt benannt: LOD0, LOD1, LOD2. LOD0 bezeichnet dabei das feinste Mesh. Unity erkennt die Benennung und richtet während des Imports automatisch LOD für das Gameobject ein.
- 3) Um die Teilmodelle korrekt zu positionieren, werden die Meshes bereits zu ihren lokalen Koordinatensystemen ausgerichtet. Dazu wird jeweils temporär ein anliegendes Teilmodell geladen, um die Modelle manuell zu diesem auszurichten.
- 4) Vor dem Export müssen die Normal-Maps bereinigt werden, falls diese noch auffällige Fehler aus dem Ursprungsmodell beinhalten. Hier wird dafür Adobe Photoshop verwendet, um mit dem Clone-Stamp-Fehler zu retuschieren.

- 5) Abschließend wird eine *.fbx* Datei exportiert, die alle Materialien und Meshes enthält.

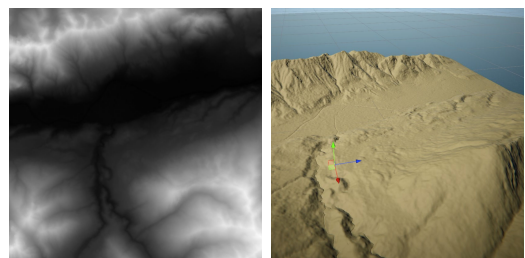
C. Gamification

In diesem Kapitel wird die Gamification in Unity beschrieben. Das beinhaltet die Erstellung einer Landschaft und die Integration der Teilmodelle aus Abschnitt II-B. Des Weiteren wird ein Third-Person-Player erstellt, der auf einem Rodel den Eiskanal befahren kann. Letztlich soll eine winterliche Landschaft entstehen.

1) *Unity Projekt*: Das Projekt wird in der Universal Render Pipeline (URP) [9] erstellt, um eine möglichst breite Auswahl an Plattformen zu unterstützen. Außerdem wird es für eine Plattformen mit durchschnittlicher Leistung ausgelegt.

Des Weiteren werden Assets von folgenden Drittanbietern verwendet: Procedural Worlds [20], Nature Manufacturer [21]. Hier werden Tools und Assets für die Erstellung der Open-World-Umgebung genutzt.

2) *World Design*: Für die Erstellung der Landschaft wird das Unity-Plugin *Gaia-Manager* [22] – ein Design-Tool für realistische Welten in Unity – verwendet. Gaia liefert viele hilfreiche Werkzeuge wie Beleuchtungen, Player und einen World-Designer, der in dieser Arbeit dazu genutzt wird, aus einer Heightmap ein realistisches Terrain der Umgebung in Innsbruck zu erstellen.

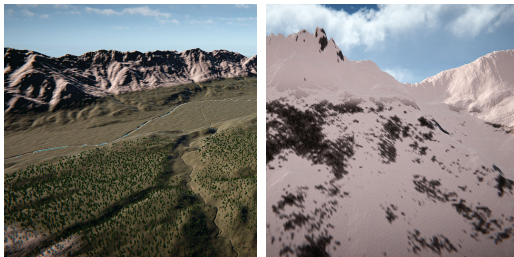


(a) Heightmap

(b) Unity Terrain

Abbildung 8: Unity Terrain und die zugehörige Graustufen-Heightmap

Um dem Terrain eine realistische Struktur und Farbe zu verleihen sowie die Umgebung mit einer Vegetation zu bevölkern, wird ein Biome-Spawner erstellt, welcher Texturen, Bäume und Sträucher erzeugt. Hier werden die Regeln für Gras-, Schnee- und Felsuntergründe erstellt, um Texturen und Prefabs an den richtigen Stellen des Terrains zu generieren. In Abbildung 9 sind die verschiedenen Texturen im Terrain zu erkennen.



(a) Terrain mit Textur

(b) Textur Detail 1



(c) Textur Detail 2

Abbildung 9: Unity Terrain und Texturen

3) *Shader*: Da die Aufnahmen der Photogrammetrie aus Kapitel II-B im Sommer gemacht wurden, ist auch die Textur der Teilmodelle sommerlich. Um eine winterliche Landschaft zu erzeugen, wird ein Shader in Unity's Shader Graph erstellt. Der Shader ist ein Programm, welches auf der GPU läuft und Einfluss auf Farbe, Position und andere Werte der Flächen eines Gameobjects nehmen kann [23]. Damit können Textur, Volumen und Farbe von Schnee auf dem Eiskanal simuliert werden. Im Abtausch dafür bindet ein Shader Ressourcen der GPU.

Shader Graph [24] ist ein grafisches Interface, in dem Shader erstellt werden können. Dies geschieht über die Erstellung von Knotenpunkten und Verbindungen.

Mithilfe des Shaders wird zum einen die Originaltextur (Base-Map und Normal-Map) der Bahn mit einer Schnee-Textur überschrieben, zum anderen wird den schneebedeckten Flächen ein zusätzliches Volumen gegeben, um eine Schneedicke zu simulieren. Außerdem darf die Textur nicht das gesamte Objekt bedecken, sondern nur die Stellen, die normalerweise mit Schnee bedeckt wären. Für einen realistischen Effekt soll an den restlichen Stellen die ursprüngliche Textur zu sehen sein und ein fließender Übergang zu den schneebedeckten Stellen stattfinden. Abbildung 10 zeigt den Effekt des Shaders.



(a) Ohne Shader



(b) Mit Shader

Abbildung 10: Snow-Shader Vergleich

4) *Laufzeit*: Für die Interaktion mit der Umgebung werden in Unity Gameobjects und eine Laufzeit mit einem Third-Person-Player erstellt. Mithilfe der

Physik-Engine [25] und individuellen Skripten für Gameobjects wird das gewünschte Verhalten definiert.

a) *Gameobjects*: Neben dem Terrain werden folgende Elemente als Gameobjects in das Projekt integriert: Der Eiskanal mit seinen 8 Teilmodellen, der Player, das Licht und die Hauptkamera (MC) – welche über Gaia implementiert sind – und der Rodel, dem das Hauptskript für die Rodelfahrt zugeordnet ist (siehe Abschnitt II-C4b). Außerdem gibt es zwei weitere Kameras, eine Followcamera (FWC) und eine Flycamera (FYC). Die FWC ist nur während der Rodelfahrt aktiv und folgt dem Fahrer. Die FYC dient dem Flug über das Gelände während einer Simulation.

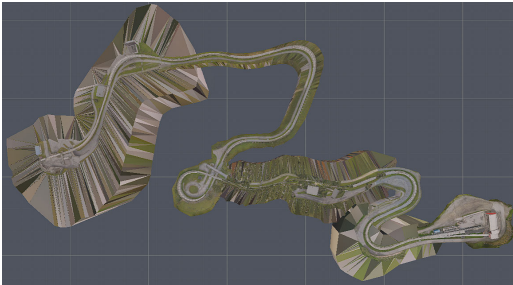


Abbildung 11: Alle Teilmodelle als Gesamtmodell in Unity

b) *Rodel-Dynamik*: Die Rodel-Dynamik wird über die Physik-Engine realisiert, also über die Simulation von physikalischen Eigenschaften wie Schwerkraft, Masse und Reibwerten, die den Rodel die geneigte Bahn nach unten gleiten lassen. Obwohl eine Simulation der realen Physik dadurch grundsätzlich möglich ist, wird das in dieser Arbeit nicht angestrebt. Stattdessen wird versucht, ein möglichst ansprechendes und flüssiges Erlebnis darzustellen. Dafür müssen einige Anpassungen in der Dynamik vorgenommen werden. Ein Grund dafür ist das stark vereinfachte Mesh, welches als Eiskanal fungiert. Das grobe Mesh und die Übergänge zwischen den Teilmodellen entsprechen nicht dem glatten und run-

den Untergrund einer Bobbahn und führen dazu, dass der Fahrer häufig aus der Bahn geworfen wird oder unerwartet stoppt. Dem kann über die Manipulation des Rodels in einem Skript entgegengewirkt werden.

Das Gameobject besteht aus dem 3D-Modell selbst und einigen leeren Objekten, welche als Positionsreferenzen für das Skript dienen. Sie können im Editor relativ zum Mesh positioniert werden und bewegen sich synchron zum gesamten Modell.

```
Sled_01
├─ sled (mesh)
├─ Seat (leeres
  Gameobject/Position)
├─ ExitPoint (leeres
  Gameobject/Position)
├─ LookAt (leeres
  Gameobject/Position)
├─ Follow (leeres
  Gameobject/Position)
└─ Back (leeres
  Gameobject/Position)
```

Dem übergeordneten Gameobject sind das Skript, ein Meshcollider und eine Rigidbody-Komponente hinzugefügt. Der Meshcollider wird außerdem mit dem *Sled.physicMaterial* verknüpft. Dieses beinhaltet die Werte für statische und dynamische Reibung sowie die Sprungkraft. Die Reibwerte werden im Skript angepasst, um das Gleiten und Stoppen des Rodels zu ermöglichen. Prinzipiell übernimmt das Skript folgende Aufgaben:

- Rodelfahrt starten
- Rodelfahrt stoppen
- Reset
- Steuerung während der Fahrt
- Seitliche Reibung
- Sprünge und Kollisionen vermeiden
- Schwerpunkt versetzen
- Zieleinfahrt

Elementar sind dabei die seitliche Reibung und

das Vermeiden von Kollisionen und Sprüngen. Um Rotation und seitliches Gleiten des Rodels zu vermeiden werden Kräfte auf das Gameobject aufgebracht, die bestimmten Bewegungen entgegenwirken. Für die Reibung wird eine einfache Gleitreibung berechnet und je nach lokaler Geschwindigkeit des Gameobjects in positive oder negative x-Richtung aufgebracht.

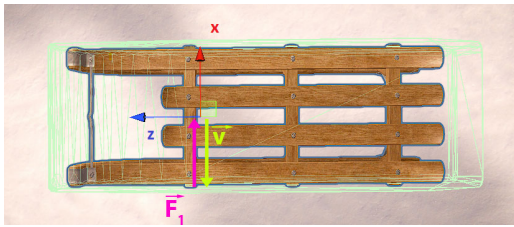


Abbildung 12: Rodel, lokale Koordinaten, Seitwärtsbewegung

Die Reibkraft wird wie folgt berechnet, mit μ als *Sideway Friction Multiplier*:

$$F_r = m \cdot g \cdot \mu. \quad (1)$$

Der Angriffspunkt liegt im lokalen Ursprung und die Richtung ist der seitlichen Bewegungsrichtung entgegenwirkend:

$$\vec{F}_1 = \begin{pmatrix} F_r \\ 0 \\ 0 \end{pmatrix} \cdot d_1, \quad (2)$$

$$\text{mit } d_1 = \begin{cases} 1, & \dot{x} < 0 \\ -1, & \dot{x} > 0 \\ 0, & \dot{x} = 0 \end{cases}. \quad (3)$$

Da diese Reibkraft im Ursprung angreift und kein Moment auf das Gameobject ausüben kann, ist eine Bewegung mit $\dot{x} \neq 0$ trotzdem möglich. Das zeigt sich in einer leicht rotierten Position des Rodels in den Kurven (siehe Abbildung 13). Dem wird wieder mit einer Gegenkraft entgegen gewirkt. Dabei wirkt die Kraft dieses mal nicht im lokalen Ursprung, sondern an der *Sled Back* Position (vgl. Abbildung

13), am hinteren Teil des Rodels. Diese Gegenkraft wirkt immer parallel zur lokalen x-Achse und ist proportional zu \dot{x} . Somit richtet diese Kraft den Rodel wieder entlang der z-Achse aus, sobald die Geschwindigkeit eine x-Komponente im Bezug auf das lokale Koordinatensystem besitzt. Die Gegenkraft berechnet sich in lokalen Koordinaten zu

$$\vec{F}_2 = \begin{pmatrix} -1 \\ 0 \\ 0 \end{pmatrix} \cdot \dot{x} \cdot C_2, \quad (4)$$

mit C_2 als *Angular Drag Multiplier*.

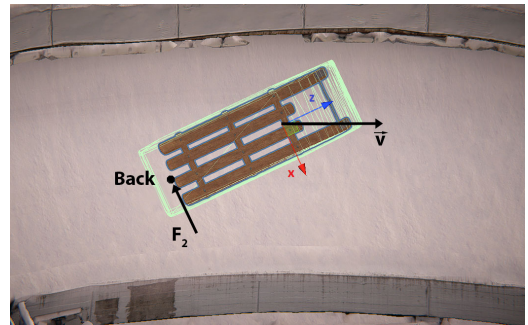


Abbildung 13: Rodel, Bewegung mit lokaler x-Komponente

Das relativ grobe Mesh der Bahn und die Übergänge zwischen den Teilmodellen sorgen oftmals für ein ungewolltes Verhalten des Rodels, welches sich durch Sprünge, Rotation oder abruptes Abbremsen bemerkbar macht. Daher werden auch hier wieder Gegenkräfte auf das Gameobject aufgebracht. Zum einen wird eine Gegenkraft F_3 in negative y-Richtung aufgebracht, welche den Rodel nach Abheben wieder auf den Untergrund drückt. Diese Kraft ist quadratisch proportional zum Abstand des Rodels vom Boden. Zum anderen wird eine Gegenkraft F_4 entlang der lokalen x-Achse aufgebracht, welche proportional zum Abstand von den Seitenwänden ist. Dadurch werden Sprünge und Berührungen mit der Seitenwand reduziert. Die beiden Kräfte berechnen sich zu:

$$\vec{F}_3 = \begin{pmatrix} 0 \\ -1 \\ 0 \end{pmatrix} \cdot D_3^2 \cdot C_3 \quad (5)$$

$$\vec{F}_4 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \cdot \left(1 - \frac{D_4}{0,3}\right) \cdot C_4 \quad (6)$$

$$D_4 = [0; 0, 4], \quad (7)$$

mit Bodenabstand D_3 , Seitenabstand D_4 , C_3 als *Vertical Force Multiplier* und C_4 als *Sideway Force Multiplier*. \vec{F}_4 soll also erst ab einem Abstand von 0,4 m oder weniger wirken. D_4 ist dabei der Abstand zur linken Seite. Die Gegenkraft zum rechten Abstand erfolgt analog.

III. ERGEBNISSE

A. Photogrammetrie

Tabelle III listet die Teilprojekte mit Berechnungszeiten (BZ) und Datenmenge der Photogrammetrie auf. Hier ist zu beachten, dass Abschnitt 1 auf hoher Qualität berechnet wurde, was sich in der Berechnungsdauer bemerkbar macht.

Tabelle III: Vergleich Berechnungsdauer pro Teilmodell

	Anzahl Bil- der	BZ Depth- Maps	BZ Mesh	BZ Tex- tur	Größe [GB]
1	1546	2d	24h:14m	0h:16m	99,7
2	1261	2h:12m	1h:54m	0h:29m	138
3	5046	6h:48m	8h:12m	1h:19m	127
4	5163	7h:23m	7h:35m	1h:22m	166
5	2205	3h:14m	2h:40m	0h:36m	77,6
6	4892	7h:14m	6h:07m	1h:14m	125
7	2582	3h:55m	3h	0h:42m	96,2
8	5902	n/a	7h:43m	1h:38m	120
	28597	≈ 79 h	≈ 61 h	≈ 7 h	≈ 950

Außerdem ist anzumerken, dass die Anzahl der Bilder diejenige Anzahl repräsentiert, welche tatsächlich in dem jeweiligen 3D-Modell zur Berechnung verwendet wurde. Die Anzahl der ursprünglich aufgenommenen Bilder im RAW-Format beträgt 35945 mit einer Größe von insgesamt 1079 GB.



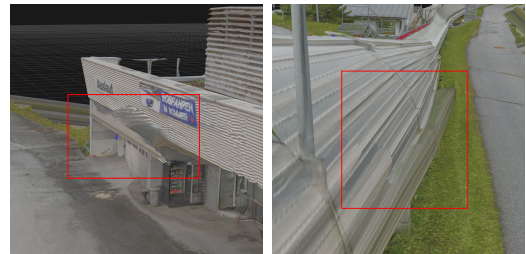
(a) Vergleich: Links 3D-Modell, rechts Foto



(b) Vergleich: Links 3D-Modell, rechts Foto

Abbildung 14: Vergleich 3D-Modell mit Foto

Abbildung 14 zeigt exemplarisch den Vergleich zwischen dem fertig texturierten 3D-Modell und dem fotografierten Objekt.



(a) Fehler im Gebäude am (b) Fehler in der letzten
Zieleinlauf Kurve

Abbildung 15: Fehler im Mesh durch fehlerhaftes Alignment

Abbildung 15 zeigt zwei Stellen, an denen eine Fehlausrichtung im Alignment-Prozess vorhanden ist. Obwohl im Rahmen dieser Arbeit die meisten solcher Defekte im Postprocessing ausgebessert wurden, sind aus Zeitgründen nicht alle Fehler beseitigt

worden.

B. Gamification

Das Projekt wurde für eine Windows-Desktop-Plattform kompiliert. Die Anwendung hat eine Größe von 15,2 GB.

Zum Testen wurde das System gemäß Tabelle II verwendet. Das Build ist auf eine Framerate von 60 FPS ausgelegt. Abbildung 16 zeigt den laufenden Spieler auf der Bahn. Die Framerate schwankt hier zwischen 40 und 60 FPS.

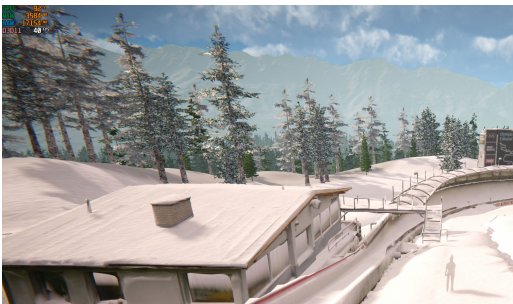


Abbildung 16: Gameplay mit Performanceanzeige

In Abbildung 17 ist eine Performanceaufzeichnung während des Gameplays zu sehen. Dabei werden ein Durchflug, eine Fahrt im Eiskanal und normale Bewegungen des Players in der Umgebung ausgeführt.

Die GPU ist dabei fast vollständig ausgelastet und läuft den Großteil der Zeit mit einer Auslastung von 90-100%. Minimalwerte, die von der GPU erreicht werden, liegen bei ca. 60% Auslastung. Die Nutzung des VRAM liegt bei bis zu 3918 MB und kratzt dabei ebenso am Limit der verwendeten Grafikkarte.

Die Auslastung des Arbeitsspeichers liegt bei bis zu 24736 MB. Dabei liegt die Vermutung nahe, dass die hohe Auslastung von RAM und VRAM durch die großen Texturen verursacht werden. Diese liegen pro Teilmodell unkomprimiert im *.jpg* Format bei bis zu 1,7 GB für LOD 0.

Die CPU-Auslastung liegt mit 5-43% im moderaten Bereich, wobei die Auslastung den größten Teil der Zeit eher im unteren Bereich liegt.

Insgesamt kann die maximale Framerate von 60 FPS während der Ausführung oft erreicht werden, sie schwankt dabei die meiste Zeit zwischen 40 und 60 FPS. Während der Animation, bei der das Gelände überflogen wird, liegt sie jedoch eher im unteren Bereich bei ca. 30 FPS und kann Minimalwerte von 17 FPS annehmen, sodass ein Stocken des Bildes auftreten kann.

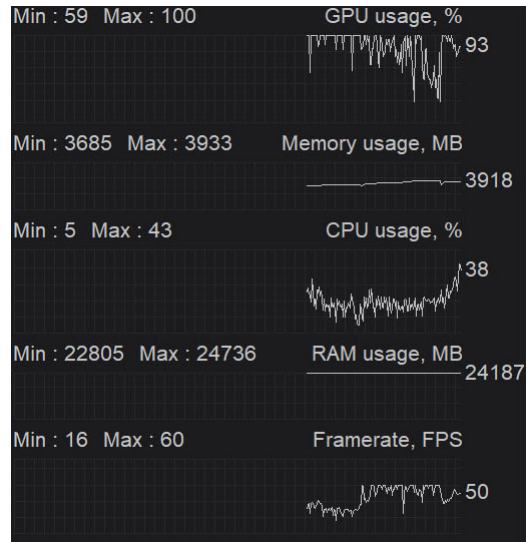


Abbildung 17: Aufzeichnung der Performance

IV. CONCLUSIO

In dieser Arbeit wurde ein Workflow entwickelt, welcher die Erstellung einer virtuellen Welt nach einem realen Vorbild beschreibt. Ferner wurde gezeigt, wie mithilfe der Photogrammetrie optimierte 3D-Modelle aus Bilddaten erstellt werden können und wie diese Modelle über Unity in eine spielbare Umgebung implementiert werden.

Anhand des Olympia-Eiskanals in Igls wurde gezeigt, dass nach diesem Workflow eine großflächige und komplexe Umgebung virtualisiert werden kann. Aus 28597 Bildern konnten 8 Teilmodelle generiert und in Unity so implementiert werden, dass die

Bahn virtuell befahren werden kann. Die produzierte Datenmenge der Source-Bilder, RC-Projekte und des Unity-Projekts beträgt dabei insgesamt ca. 2 TB.

Mit dem abschließenden Stand dieser Arbeit gibt es verschieden Möglichkeiten für weiterführende Projekte. Eines der wohl interessantesten Bereiche ist die Erstellung einer Umgebung, die mit einer VR-Brille betreten werden kann. Unity bietet dazu die nötigen Werkzeuge, um das Projekt auch für VR-Headsets kompilieren zu können. Der Großteil der weiterführenden Arbeit würde jedoch in der Optimierung und Verbesserung der aktuellen 3D-Modelle liegen, da in dem zeitlich begrenzten Rahmen dieser Masterarbeit keine vollständige Ausbesserung aller Fehler in Meshes oder Texturen stattgefunden hat. Um die Qualität des aktuellen Modells zu verbessern, müssen solche Fehler bereinigt und vor allem die Laufbahn des Eiskanal und die Übergänge zwischen den Teilmodellen so optimiert werden, dass eine flüssige und fehlerfreie Rodelfahrt möglich ist.

Weitergehend kann auch ein anderes Konzept für die Rodelfahrt angewendet werden, bei dem der Rodel einen vordefinierten Pfad entlang des Eiskanal abfährt. Diese Methode ist vor allem für Werbezwecke interessant, da unabhängig vom 3D-Modell eine sehr flüssige und stabile Bahnkurve animiert werden kann.

Für eine komplexere Nutzung wie eine Trainings-simulation für Sportler ist jedoch die Anwendung der Physik-Engine – wie es in dieser Arbeit der Fall ist – die bessere Alternative. So kann ein realistischeres Verhalten simuliert und mehr Benutzerinput definiert werden. Hier wäre es sehr interessant, mit Sportlern in Kontakt zu treten und den Bedarf an virtuellen Trainingsmöglichkeiten zu erörtern. Auf diese Weise könnte gemeinsam ein realitätsnahes Userinterface aufgebaut werden, um eine Rodelfahrt zu simulieren.

Zudem können für weitere Projekte dieser Art Methoden und Hardware in dem Prozess aus dieser Arbeit optimiert werden. So können zum Beispiel kleinere Teilmodelle erstellt werden, was mit der gleichen Datenmenge an Texturen und gleichbleibender Anzahl an Flächen pro Modell eine Steigerung

der Qualität zu Folge hat. Auch eine professionelle Kamera mit einem großen Bildsensor im Vollformat würde eine deutliche Anhebung der Modellqualität bewirken. Denn durch einen großen Sensor lassen sich einerseits leichter scharfe und kaum verrauschte Datensätze erstellen und andererseits Daten unter sehr schlechten Bedingungen aufnehmen, bei denen ein kleiner Sensor schon keine brauchbaren Bilder mehr produzieren kann. Generell lässt sich die nahe-liegende These aufstellen, dass mit einer leistungs-starken und modernen Hardware – das heißt Auf-nahmegeräte und Rechnerkomponenten – die Qualität gesteigert und die Arbeitszeit reduziert werden kann.

LITERATUR

- [1] “Olympiaworld innsbruck: Olympia eiskanal,” <https://www.olympiaworld.at/arenen-angebote/die-arenen/olympia-eiskanal/>, (Accessed on 03/14/2022).
- [2] M. Mohammadi, R. Eskola, and A. Mikkola, “Constructing a virtual environment for multibody simulation software using photogrammetry,” *Applied Sciences*, vol. 10, no. 12, p. 4079, 2020.
- [3] “Fatmap: 3d map & guides for skiing, hiking and biking,” <https://fatmap.com/around/Igls/Austria/@47.2199373,11.4298598,2942.2155001,-44.3087404,154.5891585,1076.0440186,satellite>, (Accessed on 09/20/2022).
- [4] N. Statham, “Use of photogrammetry in video games: a historical overview,” *Games and Culture*, vol. 15, no. 3, pp. 289–307, 2020.
- [5] “Unity echtzeit-entwicklungsplattform — 3d-, 2d-vr- & ar-engine,” <https://unity.com/de>, (Accessed on 04/07/2022).
- [6] “Unity - manual: Unity user manual 2020.3 (Its),” <https://docs.unity3d.com/Manual/UnityManual.html>, (Accessed on 04/07/2022).
- [7] “The most powerful real-time 3d creation tool - unreal engine,” <https://www.unrealengine.com/en-US/>, (Accessed on 04/07/2022).
- [8] “Cryengine — the complete solution for next generation game development by crytek,” <https://www.cryengine.com/>, (Accessed on 04/07/2022).
- [9] “Universal render pipeline overview — universal rp — 13.1.8,” <https://docs.unity3d.com/Packages/com.unity.render-pipelines.universal@13.1/manual/index.html>, (Accessed on 10/23/2022).

- [10] "Innsbruck-igls - google maps," <https://www.google.com/maps/place/6080+Innsbruck-Igls/@47.2218161,11.4322296,483m/data=!3m1!1e3!4m5!3m4!1s0x479d699e5cac788d:0xfb9cca39d2b08e60!8m2!3d47.23188!4d11.40854!5m1!1e4>, (Accessed on 10/15/2022).
- [11] "Unity - manual: Level of detail (lod) for meshes," <https://docs.unity3d.com/Manual/LevelOfDetail.html>, (Accessed on 10/16/2022).
- [12] "Unity - manual: Occlusion culling," <https://docs.unity3d.com/Manual/OcclusionCulling.html>, (Accessed on 10/16/2022).
- [13] A. Barnes, "Digital photogrammetry," *The Encyclopedia of Archaeological Sciences*, pp. 1–4, 2018.
- [14] R. Finsterwalder, "Photogrammetrie," in *Photogrammetrie*. de Gruyter, 2013.
- [15] W. Linder, "Digitale photogrammetrie am pc: aktuelle entwicklungen und anwendungen in den geowissenschaften," Ph.D. dissertation, Düsseldorf, Univ., Habil.-Schr., 2004, 2012.
- [16] "Dji mavic 3 bestellen - dji store," https://store.dji.com/at/product/dji-mavic-3?gclid=CjwKCAjws8yUBhAIEiwAitpEXrsjWs5xQISlbn9L5lyssylLkPhpNE_ko0KdBosUK-S8aMY2W-EB0CzSYQAvD_BwE&vid=109821, (Accessed on 05/29/2022).
- [17] "Working with components: Merging components - realitycapture support," <https://support.capturingreality.com/hc/en-us/articles/115001569011-Working-with-Components-Merging-components>, (Accessed on 10/16/2022).
- [18] "Introduction — blender manual," <https://docs.blender.org/manual/en/latest/editors/uv/introduction.html>, (Accessed on 10/20/2022).
- [19] "Bump & normal maps — blender manual," [https://docs.blender.org/manual/en/2.79/render/blender_render/textures/properties/influence/bump_normal.html#:~:text=Normal%20maps%20in%20Blender%20store,to%20Z%20\(0.0%20%2D%201.0\)](https://docs.blender.org/manual/en/2.79/render/blender_render/textures/properties/influence/bump_normal.html#:~:text=Normal%20maps%20in%20Blender%20store,to%20Z%20(0.0%20%2D%201.0)), (Accessed on 10/20/2022).
- [20] "Procedural worlds: Content creation & delivery tools," <https://www.procedural-worlds.com/>, (Accessed on 10/23/2022).
- [21] "Home — nature manufacture," <https://naturemanufacture.com/#:~:text=Nature%20Manufacture%20is%20a%20studio,the%20needs%20of%20the%20audience.>, (Accessed on 10/23/2022).
- [22] "Gaia — procedural worlds," <https://www.procedural-worlds.com/products/indie/gaia/>, (Accessed on 05/30/2022).
- [23] "Unity - manual: Shaders introduction," <https://docs.unity3d.com/Manual/shader-introduction.html>, (Accessed on 10/21/2022).
- [24] "Unity - manual: Using shader graph," <https://docs.unity3d.com/Manual/shader-graph.html>, (Accessed on 10/22/2022).
- [25] "Unity - manual: Physics," <https://docs.unity3d.com/Manual/PhysicsSection.html>, (Accessed on 10/21/2022).

Dominic Ludwig ist Student im Bereich Mechatronik am MCI Innsbruck. Neben seiner Vertiefung im Maschinenbau ist er tätig als Softwareentwickler bei web-crossing GmbH.

Concept Development for the Rotor Assembly Line of an Interior Permanent Magnet Synchronous Motor (IPMSM)

Daniel Martino, Dr. techn. Franz-Josef Falkner (supervisor), and Dipl.-Ing. Martin Schwarz (supervisor)

Abstract—This paper presents a possible process flow for the production of the rotor of an Interior Permanent Magnet Synchronous Motor (IPMSM). It is feasible for a rotor assembly line which has to ensure scalability in terms of production per year and producibility of different variants. The fixation of permanent magnets in rotor stacks has been investigated, and a state-of-the-art process, Resin Transfer Moulding (RTM), has been proposed to replace gluing. An efficient approach to integrate the RTM process is shown. Additionally, the thesis discusses the most suitable method of thermal process for joining the stack package on the rotor shaft and as well heating the rotor to reach the required temperature for fixating the magnets by RTM. In addition, the paper proposes a method of how the RTM process can be fulfilled with the so-called Nut-and-Shoulder variant plus already joined shaft on the stack package, by launching a balancing ring design with built-in ventilation.

Index Terms—Interior Permanent Magnet Synchronous Motor (IPMSM), Rotor Assembly, Resin Transfer Moulding, Process Flow

I. INTRODUCTION

IN the manufacture of Interior Permanent Magnet Synchronous Motors (IPMSMs), efficient planning of the entire process flow is essential to ensure optimum performance and reliability of the motor. The rotor, a key component of the motor, is primarily made up of a shaft, stacks consisting of coated electrical sheets connected by an interlock connection, permanent magnets fixed in the stacks, and balancing discs. In this paper, a possible process

flow is presented and investigated. This is significantly influenced by an introduced state-of-the-art magnet fixing technique, RTM. In this process, a thermoset is injected into cavities under pressure and at $T \approx 175^\circ\text{C}$, resulting in a high filling factor. Previously, push-out tests were conducted at the same company to evaluate the performance of different magnet fixation methods after several heating and cooling cycles. It was found that magnets fixed by RTM consistently showed about the same force requirement to be pushed out of the stack, even after several hundred cycles. Glued stacks, on the other hand, showed strongly decreasing ejection forces with an increasing number of cycles. In addition, the gluing process generally leads to inconsistencies in the amount of glue between the magnets and the stacks.

The integration of RTM influences the entire process flow and, consequently, the entire series production line. This paper aims to present and investigate an optimised process flow in the rotor assembly line, considering the introduction of the RTM process for magnet fixation.

Furthermore, the system should be able to produce two design variants, the Standard variant in which the balancing discs are press-fitted onto the shaft on both sides of the stack package and a so-called Nut-and-Shoulder variant. In this variant, a shoulder is provided on the shaft on one side of the fitting surface and a thread on the other side. The stack package with balancing discs is pressed onto the shoulder

with a specific force and this pretension on the stack package is preserved by tightening a groove nut.

In addition, this paper addresses the implementation of both design variants in one system.

II. METHODS

A. First Decisions Taken for Process Flow

The first decision was that the stacks should not be moulded individually (single-stack RTM) but as a composite (multi-stack RTM). As Single-stack RTM is not possible within a reasonable cycle time without making a very high investment in many RTM machines and taking up a huge area of the production facility. As the main reason, the multi-stack RTM increases the stiffness of the entire rotor which leads to a higher quality, especially in terms of NVH. The stiffening effect should be created by the connections from the thermoset, which are built up in the magnetic cavities between the stacks during RTM. The second essential question, which strongly influences the process and thus the assembly line, was the following: Should the RTM be done with or without a joined shaft in the stack package? The decision was made in favour of RTM with an already joined shaft. As the connections between the stacks made out of the moulded thermoset are retained as well as possible with this method. If RTM was carried out before joining the shaft, the stacks would be first aligned on a mandrel, and then after RTM, the shaft could be joined with the stack package. After the joining, a heat transfer between the heated stack package and cold shaft would take place immediately, thus the clearance fit between stacks and shaft that occurs due to the temperature difference would become an interference fit. This would always lead to a relative movement between the stacks. Although this could be kept to a minimum with a great deal of process effort, in most cases, the thin connecting layers, made of the thermoset would experience cracks or even break.

Most of the components fed into the system are scanned using data matrix codes (DMCs), which are

2-dimensional codes that can be scanned and hence, data can be stored and accessed via them. A distinction is made whether only the batch can be assigned to certain parts or each component individually. This is crucial, as the parameters and other information recorded during the process can be traced back later.

B. Stacks and Magnets Piling

The actual process flow begins at this point. The rotor assembly line must be continuously loaded with three different stack variants, which only differ in the angle between the key and the cavity profile, to generate the so-called skewing angle. Between each stack, the magnets of the stack package are positioned offset by a certain angle [1]. This angle is used to reduce the so-called cogging torque which has negative effects on the performance [2]. Stacks are positioned on dummy shafts, situated either on a turntable or a conveyor moving in a loop. Once a stack is in place on the shaft, magnets are inserted into the cavities of the stack. Subsequently, the next stack is positioned atop the previously filled stack and equipped with magnets. The magnets are inserted into the cavities using several SCARA robots. The magnets are fed into the system with so-called C-bars, which contain the magnets. Another approach involves using bulk material, where the magnets are poured into the rotor assembly line bit by bit. The permanent magnets are sintered and brittle. Thus, there would be a higher risk of damaged magnets, and so the other variant was chosen.

C. Heating Processes

To avoid any kind of damage, the stack package should be joined on the shaft force less. This can be achieved by creating a significant temperature difference between the two components, resulting in a clearance fit.

To obtain an approximate relationship between the fitting range as a function of the temperature difference, between the cold shaft and the warm stacks, the coefficient of thermal expansion α was used, which is defined as

[3]

$$\alpha = \frac{\Delta L}{L_0 \Delta T}. \quad (1)$$

In Eqn. (1) ΔL is the difference in diameter caused by the temperature difference ΔT and L_0 is the initial diameter.

Since a homogeneous temperature of $\approx 175^\circ\text{C}$ is required for RTM, cooling the shaft does not make sense, as residual heat inside the rotor is needed for the process. Therefore, only the stack package is being heated. This initial heating of the stack package can be carried out either in an oven or through induction. The same applies to the second heating process step, in which the composite of stack packages and shaft must be brought to a homogenised temperature. For this purpose, both heating methods employed in each of the two thermal processes were thoroughly investigated through a combination of experiments, calculations, and simulations.

To make a correct decision for the heating methodology, but also to choose the correct parameters and boundary conditions in case of oven heating, the aim was to create a functional and very realistic FEM simulation. To validate it, an experiment was carried out in reality. Based on it, various parameters could be optimised. The process is illustrated in Fig. 1 and Fig. 2.

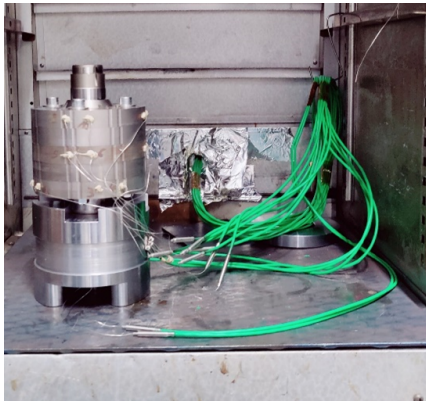


Fig. 1. Experiment in real oven.

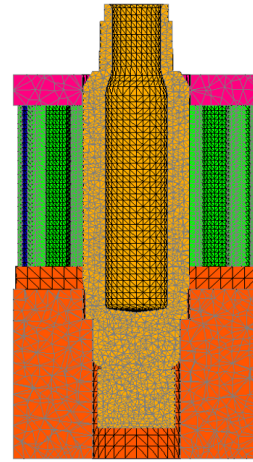


Fig. 2. The Mesh was prepared with the software Altair 2021 Simlab and consisted of quadratic TETRA10 elements. The FEM simulation was performed with the software Marc Mentat 2022.

An oven for the second heating step could lead to a critical surface hardness reduction, as the shafts are made of steel with a low carbon content and since a high surface hardness is required on the helical gearing of the shaft, they need to be case-hardened. During tempering (sub-process of case-hardening), steel gets heated between 160°C and 200°C for 2 h to 4 h [4]. As the rotor shaft is exposed to similar temperatures during heating for RTM, the loss of hardness was examined. Several shafts were employed, which were composed of the same material as the rotor shaft. First, part of the gearing was removed, followed by a meticulous polishing process to evaluate the surface hardness by applying the HV1 method. Hardness measurements were then taken at ten points along the tooth flank surface, precisely 0.1, mm from the border. The median of these measurements was used as a representative value for the overall surface hardness. Then some parts were heated at 180°C and others at 190°C . The dwell time was set 30% higher than the maximum time the part should normally spend in the oven. The parts

were then trimmed again on the gearing to determine the surface hardness in the same manner as before.

To check whether the two heating process steps could be realised in a reasonable cycle time by an induction heating process without damaging the stacks, and to find the optimal parameters for this process, various test runs have been carried out with different parameters, as can be seen in Fig. 3. A further target was to understand whether only an external coil would be sufficient or if an external coil would be required in conjunction with a coil from the inside.

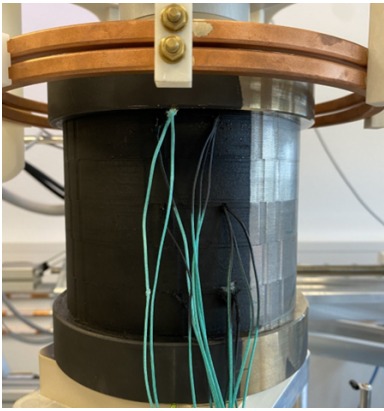


Fig. 3. Prototype inductive heating.

D. Resin Transfer Moulding (RTM)

On top of the hot rotor, the so-called runner gate plate is placed at about the same temperature. This is a tooling that is precisely aligned with the top stack and is equipped with channels, through which the thermoset can be injected into the rotor.

Once the runner gate plate is on the rotor, it is placed in the RTM machine and the stack package is compressed with a force of approximately 200 kN to guarantee absolute tightness. Therefore, compression of the stacks, whose laminations are connected by an interlock connection, is to be expected. Hence, it was important to understand how much the stacks are

compressed and whether this could cause damage to the magnets. For this purpose, an experiment was carried out in which different stacks were loaded with a hydraulic press with different forces of up to 200 kN. The height difference was measured and the relative compression was calculated from this.

The thermoset is fed into the machine in form of small pellets ($\varnothing \approx 20$ mm). These are preheated in the piston's displacement chamber, which then presses them through the runner gate plate into the rotor. The properties of the material ensure that it becomes very fluid under high temperatures and pressure.

During the piston's extension, the pressure on the thermoset increases. After injection, the curing process starts, which lasts six times as long as the injection, making it the bottleneck for the cycle time of this process. According to the thermoset supplier, the rotor's temperature is decisive for the curing time and increases exponentially with a temperature reduction. Prototypes were moulded at a slightly lower temperature because the curing time was adjusted accordingly.

It should be noted that the bottom of the tooling that clamps the stack package has vent holes built into it. The air in the magnetic cavities escapes through these, preventing air entrapments, which could have negative effects on the final product. These holes should not exceed a diameter of $50 \mu\text{m}$ because, according to the supplier, the thermoset compound used in this system, although liquid, hardens in very small cavities and thus cannot escape through them. In general, the polymer used for RTM is expected to cure at a certain cavity height [5].

Since the tooling's vent cannot be used with the Nut-and-Shoulder variant, as the end ring must already be placed on the shaft before RTM, a design solution had to be found for this problem as well.

After the curing process has been completed, the runner gate plate is removed from the rotor. In the process, the sprues must be broken off by a rotating movement, as this breaks them off much more cleanly. After the runner gate plate has been

removed from the rotor, it is cleaned in a specially designed machine.

E. Cooling

The rotor is brushed on the upper side of the stack package to refine the surface at the break-off points of the sprue. Afterwards, it is placed into a cooling tunnel with strong ventilation to increase the forced convection and therefore speed up the process.

F. Balancing Disc attachment

- For the Standard variant, two discs must be pressed on the shaft with a force, as an interference fit has been designed between the shaft and the disc to prevent twisting, which is essential for the function of the balancing.
- For the Nut-and-Shoulder variant, only one disc is applied, because the lower one is already on the shaft between the shoulder and the stack package. This is done without force, as no interference fit is required to prevent twisting after balancing. This is prevented by the load applied later on the stack package as well as by an aligning longitudinal key on the shaft.

After this step, for the Nut-and-Shoulder variant, a force is applied to the stack package through a hydraulic press. The resulting tension on the stack package is retained by tightening a groove nut. The exact specifications for force and tightening torque were determined by performing a modal analysis. Attention was paid to the combination of force and torque, at which the natural frequencies and thus the stiffness of the composite is highest.

G. Balancing

Balancing is carried out according to ISO 1940-1, whereby a balance quality grade of G 2.5 is to be achieved, which according to the standard is suitable for generators and electric motors that exceed a shaft height of 80 mm and a speed of 950 rpm [6]. Since a quality class of G 2.5 is to be achieved, the maximum

speed during the operation of the motor as well as the weight of the rotor shaft are known, and the maximum permissible unbalance can be extracted from a diagram in [6].

If the unbalance is higher during the unbalance measurement, the rotor is balanced. The mass is reduced selectively by making holes at certain points on the end discs. This process is repeated until the rotor no longer exceeds the specified maximum admissible unbalance.

III. RESULTS

A. Heating Processes

Through the described calculation, it was possible to determine to which temperature the stack package must be heated (see Fig. 4).

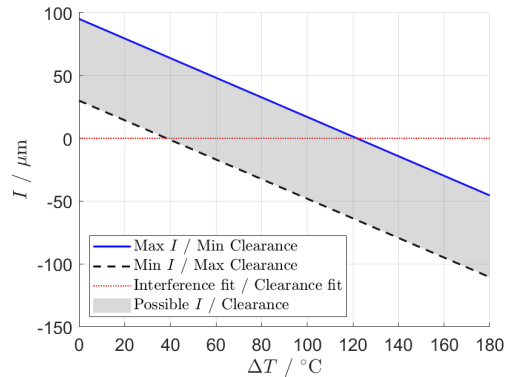


Fig. 4. The coloured area indicates the range in which the interference I /clearance can be located depending on the exact diameter of the components (within the tolerances) and the temperature difference ΔT .

In Fig. 4, it can be observed that an interference of 0 is achieved with the configuration of dimensions providing the maximum allowable interference within the specified tolerances at $\Delta T \approx 125^\circ\text{C}$. Therefore, the stack package should be heated to at least 150°C to join the components without force. Since a higher clearance is desirable for handling,

the exact temperature of the shaft can vary greatly, and high residual heat is advantageous for the further course of the process flow, heating of the stack package to 180 °C is targeted.

Since coil heating worked well without damaging the stack package in the experiment and only took up a fraction of the time of the oven, as well as an advantageous smaller size of the machines, induction was chosen for this step. Fig. 5 compares heating with only one external coil and heating with two coils.

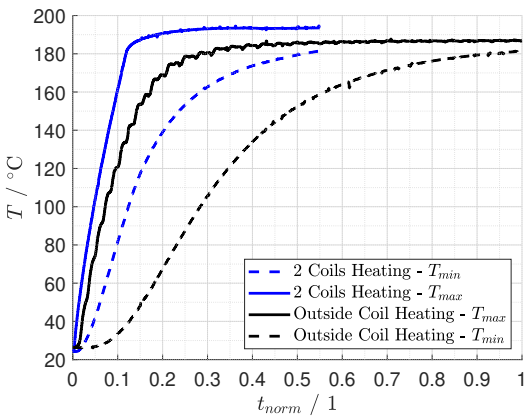


Fig. 5. The maximum temperature difference within the stack package and the required heating time are both smaller with two coils.

To avoid an excessive number of induction machines but also to achieve more uniform heating of the stack package, an inner coil in combination with an outer coil has to be used.

To carry out the second heating process for the RTM by an induction system, an internal as well as an external inductor would have to be utilised. As we had opted for RTM with a joined shaft, an internal inductor would have to be inserted into the bore of the shaft. As in the experiments performed it was essential that both the outer and inner coils were controlled. The decision was taken against a control-free internal heating system, because in series operation,

depending on the season and storage locations, we have to reckon with shafts whose temperature can vary between 0 °C and 60 °C.

The internal regulation employing a pyrometer requires a certain amount of space, which currently no supplier has been able to undercut. To avoid compromising flexibility or necessitating modifications to pre-planned projects, the decision was made in favour of the oven solution.

To determine the optimal boundary conditions, various parameters were investigated. The most important one was the oven temperature, as can be seen in Fig. 7.

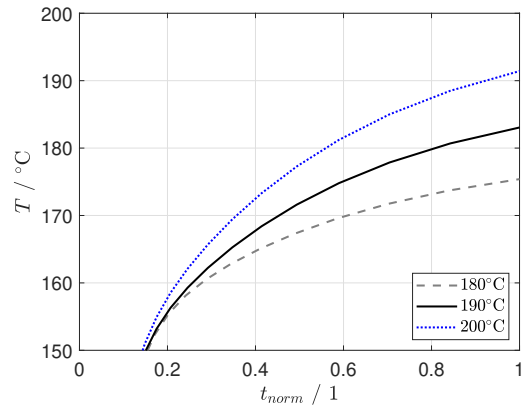


Fig. 6. Temperature Progression of the helical gearing at different oven temperatures.

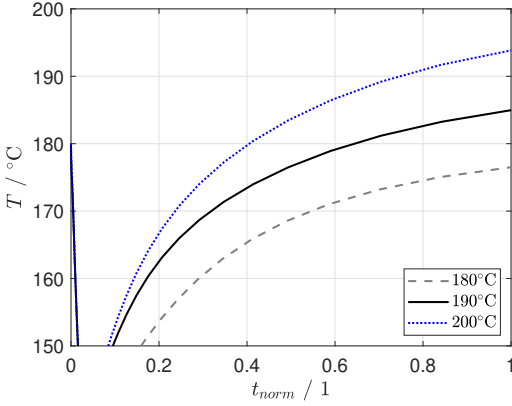


Fig. 7. Temperature Progression of the stacks at different oven temperatures.

Since a relatively homogenised temperature of the entire rotor of about 175°C has to be achieved, a clever dwell-time in the oven was chosen.

For the investigation of the surface hardness loss measurement, Fig. 9 was created.

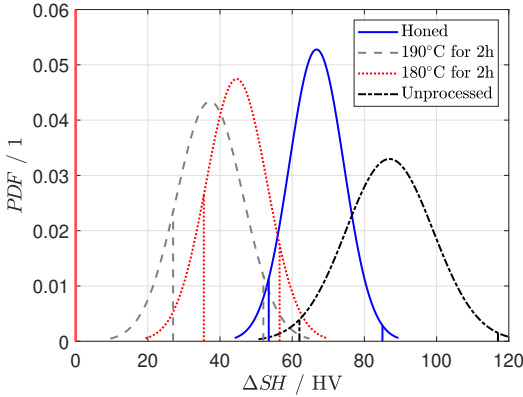


Fig. 8. Probability density function PDF of Surface Hardness above the tolerance limit ΔSH . The effective minimum and maximum values measured for each bell curve are indicated by the vertical lines.

The mean and standard deviation were determined according to the three-sigma system. According to

this, about 99.7% of a normal distribution lies within three times the standard deviation of the mean [7]. Therefore, a reduction in surface hardness will happen, but not to a critical extent.

B. RTM

The findings of the stack compression investigation (Section II-D) are shown in Fig. 9.

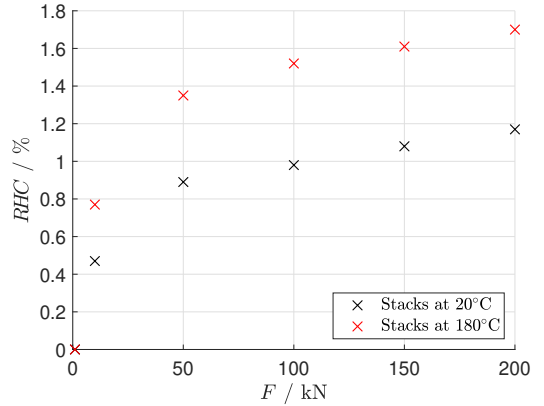


Fig. 9. Relative height compression RHC .

The temperature of the stack has a big impact on the compression. Since a temperature of $\approx 180^{\circ}\text{C}$ of the stacks is to be expected during RTM, this curve is more representative. This diagram shows that a compression of the stacks of up to 1.8% can be expected. It can also be seen that at a force of 50 kN, almost 1.5% compression is already achieved.

A compression of about 1.8% of the stack could be problematic, having a minimum difference between stack height and magnet height of about 0.5% of the stack height. So there was a concrete risk that the whole force could act on a magnet and since it is sintered, it would most likely be damaged. As a solution to this, it was foreseen that in the future the supplier must ensure, that the stacks lie within the tolerances, while a force of 50kN is acting on them.

As mentioned in Section II-D, the balancing discs were redesigned to ensure ventilation for a properly

working RTM process. Instead of using holes, which would have required expensive reworking, a design was developed that uses so-called venting grooves, which can be cast directly with the disc. Based on successful trials in the production of moulded prototypes, the design proved its effectiveness, so a patent application was filed.

IV. CONCLUSION

With this paper, a realisable process flow could be planned, which is feasible for both design variants and includes the state-of-the-art RTM magnet fixing method. This has been proved by several performed tests, experiments, simulations, and most importantly the produced prototypes which demonstrated high quality. Nevertheless, several possible improvements are still being examined. An example of a concrete idea that is already being worked on and tested is the introduction of small notches that are cleverly punched into some laminations within the magnet cavity of the stacks. The magnets would have to be pressed into the slot with a slight force, but afterwards, the handling of the rotors before the RTM process would be easier, as there would no longer be a need to ensure that the magnets had to be held.

A further step would be to separate the shaft into two parts, by connecting the helical gearing through a spline connection at the end of the rotor assembly line. This would avoid the costly and very energy-consuming case-hardening process for the whole shaft, as this would only be done for the helical gearing. Furthermore, this would exclude the negative surface hardness-losing effect which occurs during case-hardening.

ACKNOWLEDGMENT

The author would like to thank his supervisors Dr. Franz-Josef Falkner (MCI) and Dipl.-Ing. Martin Schwarz (GKN) for their guidance, the entire manufacturing engineering team for its support, and also the company GKN Automotive, which made this very interesting project possible in the first place.

REFERENCES

- [1] R. Islam, I. Husain, A. Fardoun, and K. McLaughlin, "Permanent-magnet synchronous motor magnet designs with skewing for torque ripple and cogging torque reduction," *IEEE Transactions on Industry Applications*, vol. 45, no. 1, pp. 152–160, 2009.
- [2] C. Studer, A. Keyhani, T. Sebastian, and S. Murthy, "Study of cogging torque in permanent magnet machines," in *IAS '97. Conference Record of the 1997 IEEE Industry Applications Conference Thirty-Second IAS Annual Meeting*, vol. 1, 1997, pp. 42–49 vol.1.
- [3] "Chapter 3 - intrinsic voids in polymers," in *Voids in Materials*, G. M. Gladysz and K. K. Chawla, Eds. Amsterdam: Elsevier, 2015, pp. 37–48.
- [4] V. Läßle, *Wärmebehandlung des Stahls: Grundlagen, Verfahren und Werkstoffe*, 8th ed. Verlag Europa-Lehrmittel, 2003.
- [5] "Simulation and experimental analysis of large area substrate overmolding with epoxy molding compounds," *Microelectronics Reliability*, vol. 51, no. 3, pp. 668–675, 2011.
- [6] International Organization for Standardization, "Mechanical Vibration – Balance Quality Requirements for Rotors in a Constant (Rigid) State – Part 1: Specification and Verification of Balance Tolerances," 2003, ISO 1940-1:2003.
- [7] E. M. White and R. Schroeder, "A simultaneous control chart," *Journal of Quality Technology*, vol. 19, no. 1, pp. 1–10, 1987.

Daniel Martino Student of the Master's degree programme Mechatronics & Smart Technologies at MCI in Innsbruck/Austria and eMotor Manufacturing Engineer at GKN Automotive in Bruneck/Italy



Experimentelle Bestimmung der Biegesteifigkeit eines Alpinskis

Florian Mikula, Franz-Josef Falkner (Betreuer)

Kurzfassung—Die Performance eines Alpinskis ist von vielen Parametern abhängig. Unter anderem beeinflusst die Biegesteifigkeit das Fahrverhalten eines Skis entscheidend. Durch den komplexen Aufbau eines Skis ist die Biegesteifigkeit nur schwer zu berechnen. Die Kenntnis über den Verlauf der Steifigkeit würde den Entwicklungsprozess eines Skis allerdings beschleunigen und verbessern. Dieser Verlauf kann jedoch nicht direkt gemessen werden.

Diese Arbeit beschreibt die Entwicklung eines Verfahrens zur Ermittlung der Biegesteifigkeit aus einer gemessenen Biegelinie. Das Verfahren basiert auf der Euler-Bernoulli-Balkentheorie. Mathematisch führt die Bestimmung der Biegesteifigkeit zu einem schlecht formulierten, inversen Problem. Dieses wird über die Optimierung der kleinsten Fehlerquadrate gelöst. Um ein Oszillieren des Optimierungsverfahrens zu verhindern, wird eine Regularisierung basierend auf der Tikhonov-Regularisierung angewendet.

Die Validität des Verfahrens wird mit Messobjekten aus Stahl, von welchen die Biegesteifigkeit bekannt ist, aufgezeigt. Abschließend wird das Verfahren für einen Alpinski angewendet.

Schlagwörter—Ermittlung Biegesteifigkeit, Inverses Problem, Tikhonov Regularisierung, Biegesteifigkeit Alpinski

I. EINLEITUNG

DER Verlauf der Biegesteifigkeit eines Alpinskis ist bei der Entwicklung eines Skis äußerst relevant. Mithilfe eines Prüfstandes und einer entsprechenden Auswertung der Messergebnisse kann die veränderliche Biegesteifigkeit eruiert

F. Mikula studiert am MCI, Innsbruck, Österreich, e-mail: mf0675@mci4me.at.

F. Falkner unterrichtet am MCI, Innsbruck, Österreich. Manuskript eingereicht am 30. September 2022.

werden. Ein solcher Prüfstand wurde in einer vorausgehenden Masterarbeit [1] entwickelt und gefertigt. Diese Arbeit beschäftigt sich mit der Bestimmung der Biegesteifigkeit aus den Messergebnissen.

Basierend auf der Euler-Bernoulli Balkentheorie wird die Durchbiegung eines Balkens $w(x)$ mit

$$\frac{\partial^2}{\partial x^2} \left(EI(x) \frac{\partial^2 w(x)}{\partial x^2} \right) = q(x) \quad (1)$$

beschrieben [2]. In der klassischen Balkentheorie wird die Verschiebung $w(x)$ mit einer gegebenen Belastung $q(x)$ und gegebener Biegesteifigkeit $EI(x)$ berechnet. Die Differentialgleichung kann mit entsprechenden Randbedingungen nach $w(x)$ gelöst werden.

Um die Biegesteifigkeit zu ermitteln, muss die Gleichung allerdings nach $EI(x)$ gelöst werden. Dazu wird die Krümmung $\partial^2 w / \partial x^2$ benötigt, welche nicht direkt gemessen werden kann. Entsprechend werden mit einem experimentellen Aufbau die Belastung und die Biegelinie ermittelt. Die Lösung für die unbekannte $EI(x)$ aus den Messergebnissen stellt dabei eine besondere Herausforderung dar. [3]

A. Vorausgehende Arbeiten

Zu erwähnen sind drei Kategorien an vorliegenden Untersuchungen:

Vorausgehende Masterarbeit: Das Fundament dieser Arbeit wurde von Hannes Schwendinger in [1] gelegt. In der Masterarbeit konnte zwar ein Verlauf der Steifigkeit über die Länge ermittelt werden, die Ergebnisse weisen allerdings Instabilitäten auf.

Skiprüfstand - Stand der Technik: Die Messung der Steifigkeitsparameter eines Skis wurde in [4] und [5] bereits untersucht. Ein Verfahren zur Ermittlung der Biege- und Torsionssteifigkeit aus einer kombinierten Belastung wurde in [4] entwickelt. Mit dem Prüfstand wird die Neigung der Biegelinie respektive der Winkel der Verschiebung gemessen. Dieses kombinierte Verfahren aus Biege- und Torsionsbelastung bringt allerdings Probleme in der Genauigkeit und Dauer der Messung mit sich.

In [5] wird ein optisches Messverfahren für die Biege- und Torsionsbelastung untersucht. Aufgrund der geringen Wiederholgenauigkeit von ± 1 mm und der geringen Anzahl der Messpunkte können allerdings keine zuverlässigen Ergebnisse erzielt werden. Der Ansatz eines optischen Verfahrens zur Messung der Torsion könnte allerdings sehr genaue Ergebnisse liefern.

Verfahren zur Ermittlung der Biegesteifigkeit: Das Auswerteverfahren ist mathematisch eine komplexe Aufgabenstellung. Die Ermittlung der Biegesteifigkeit stellt ein schlecht gestelltes, inverses Problem dar. Die Lösung solcher Probleme ist in der Mathematik von großem Interesse. [3] Mithilfe der Variationsrechnung kann das Problem in ein gut gestelltes Variationsproblem überführt werden. Mit [6] wird aufgezeigt, dass die numerische Lösung der Variationsrechnung valide Ergebnisse liefert. Dieses Prinzip wurde zur Auswertung der Biegesteifigkeit in [7] verwendet. Mit der Kombination der Variationsrechnung und der Tikhonov-Regularisierung kann das Problem numerisch effizient und stabil gelöst werden.

B. Aufbau der Arbeit

Das Messverfahren und die Modifikationen des Prüfstandes sind in Kapitel II-A dokumentiert. Die Methoden zur Ermittlung der Biegesteifigkeit werden in Kapitel II-B beschrieben. Die erreichten Ergebnisse mit den beschriebenen Methoden werden in Kapitel III dargestellt und in Kapitel IV diskutiert. Die Arbeit endet mit der Conclusio V und der Danksagung.

II. METHODEN

A. Prüfstand und Messaufbau

Der Ski-Prüfstand wird in [1] detailliert beschrieben und ist in Abbildung 1 dargestellt. In diesem Kapitel wird der Ablauf der Messung und die Datenauswertung aus [1] zusammengefasst und beschrieben.

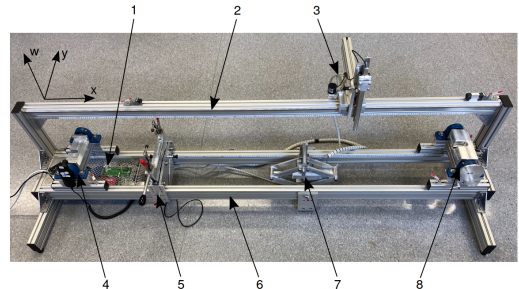


Abbildung 1. Übersicht Skiprüfstand (Stand: 2021) [1]: 1 Datenerfassung, 2 Horizontale Linearführung, 3 Messkopf, 4 Loslager, 5 Torsionsmodul (nicht verwendet), 6 Grundgestell, 7 Biegemodul, 8 Festlager

1) *Ablauf der Messung:* Bei der Messung der Biegelinie mit dem Prüfstand aus Abbildung 1 wird das Testobjekt am Festlager (8) fixiert und unter dem Loslager (4) positioniert. Mit dem Biegemodul (7) wird der Gegenstand deformiert. Die Messwerte können direkt über die Datenerfassung (1) von *National Instruments* in MATLAB eingelesen werden. Der Messkopf (3) ist mit einer vertikalen Linearführung und einer Positionsmessung ausgestattet. Beim Messvorgang wird der Messkopf horizontal (2) über die Länge des Messobjektes bewegt. Dabei werden die auftretende Kraft, die horizontale und die vertikale Position gemessen. Um das Eigengewicht des Testobjektes aus den Messergebnissen zu eliminieren, werden zwei Deformationszustände gemessen und voneinander subtrahiert.

2) *Linearitätsnachweis:* Die Gleichung (1) ist nur für schubstarre Balken und kleine Neigungen $\partial w / \partial x \ll 1$ gültig [2]. Um die Biegesteifigkeit zu ermitteln, muss der Ski bei der Messung allerdings

stark deformiert werden. Damit die Euler-Bernoulli-Balkentheorie für die großen Verschiebungen angewendet werden kann, wird das lineare Verhalten mit einer Messung überprüft.

Die Messergebnisse sind in Abbildung 2 dargestellt. Die Messung zeigt auf, dass auch bei großen Verschiebungen ein lineares Verhalten zwischen Belastung und Verformung auftritt. Damit ist die Verwendung der mechanischen Gesetze (1) zulässig.

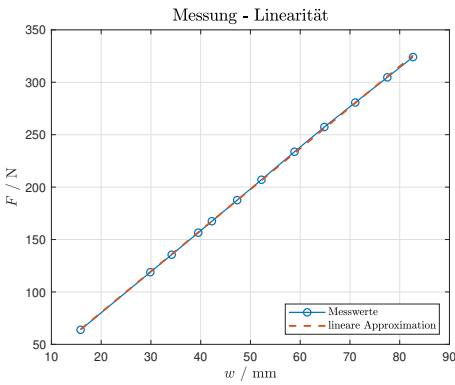


Abbildung 2. Messung des linearen Verhaltens der Deformation bei unterschiedlicher Belastung. Dazu wird ein Alpinski mit dem Prüfstand unterschiedlich groß deformiert. Dabei wird die auftretende Kraft gemessen. Zum Vergleich wird die lineare Approximation in der Abbildung dargestellt.

3) *Messgenauigkeit*: Für die Ermittlung der Messgenauigkeit des Prüfstandes wird die Standardabweichung und die maximale Abweichung einer Messreihe berechnet. Die Biegelinie des deformierten Skis wird $N = 9$ mal durchgeführt, ohne die Belastung zu ändern. Die Standardabweichung der gemessenen Biegelinien

$$\sigma_w = \sqrt{\frac{1}{N-1} \sum_{i=1}^N |w_i - \bar{w}|^2} \quad (2)$$

beträgt

$$\bar{\sigma}_w = 9.5 \mu\text{m}.$$

Die Differenz des maximalen und minimalen Wertes der Messreihe w_i an jeder Stelle x wird mit

$$\Delta w = \max(w_i) - \min(w_i) \quad (3)$$

berechnet. Der Mittelwert des Fehlers beträgt

$$\overline{\Delta w} = 29.5 \mu\text{m}.$$

B. Ermittlung der Biegesteifigkeit

Zur Ermittlung der Biegesteifigkeit wird die Gleichung (1) mit dem Biegemoment

$$\frac{\partial^2 M(x)}{\partial x^2} = q(x) \quad (4)$$

als

$$w_{,xx}(x) = -\frac{M(x)}{EI(x)} \quad (5)$$

formuliert [2]. Dabei wird die Krümmung mit $w_{,xx}(x) = \partial^2 w(x)/\partial x^2$ und die Biegesteifigkeit mit $EI(x)$ notiert. Zum Lösen der Differentialgleichung werden die Randbedingungen für den Prüfstand

$$\begin{aligned} w(x_A) &= w_A; & w_{,xx}(0) &= 0 \\ w(x_B) &= w_B; & w_{,xx}(L) &= 0 \end{aligned} \quad (6)$$

benötigt. Wie bereits in Kapitel I erwähnt, wird zur direkten Lösung der Gleichung (5) nach $EI(x)$ die Krümmung $w_{,xx}(x)$ und das Moment $M(x)$ benötigt. Mit dem Prüfstand kann allerdings nur die auftretende Kraft \hat{F} und die Biegelinie $\hat{w}(x)$ gemessen werden.

Deshalb wird die Biegesteifigkeit wie in [8] über ein Optimierungsverfahren mit einer iterativen nicht-linearen Regularisierung ermittelt. Mit dem Optimierungsverfahren wird eine Testfunktion $EI^*(x)$ mit dem Algorithmus solange angepasst, bis die optimierte Biegelinie $w^*(x)$ mit der gemessenen Biegelinie $\hat{w}(x)$ bestmöglich übereinstimmt.

1) *Optimierungsverfahren*: Für die Approximation wird das Optimierungsverfahren `lsqnonlin` in MATLAB verwendet. Die Optimierung wird in zwei Schritten durchgeführt. Zuerst wird die Biegesteifigkeit EI^* als konstant angenommen und mit dem Algorithmus optimiert. Diese Lösung dient als

Schätzung für den Initialwert EI_0 der eigentlichen Optimierung.

Für beide Optimierungsverfahren wird zur Ermittlung von $EI^*(x)$ die Zielfunktion \mathbf{J} mit dem Funktional \mathbf{F} minimiert

$$\min_{EI^*} \mathbf{J}(EI^*) = \min_{EI^*} \|\mathbf{F}(x)\|_2^2. \quad (7)$$

Die Dimensionen der Vektoren und Matrizen sind kritisch für die Lösung des Optimierungsverfahrens. Entsprechend werden im Nachfolgenden die Matrixgrößen in die Notation mit aufgenommen.

Die Lösung des Randwertproblems wird an $j = 100$ Stellen ermittelt. Das Ergebnis $\mathbf{w}_{1 \times j}^*$ entspricht der diskreten Biegelinie der aktuellen Testfunktion $EI^*(x)$.

Um das Funktional \mathbf{F} der Zielfunktion (7) zu bilden, wird die gemessene Biegelinie $\hat{w}(x)$ an j Stellen extrahiert. Aus den zwei diskreten Biegelinien wird das Funktional

$$\mathbf{F}_{\mathbf{w}, 1 \times j}(x, EI^*) = \mathbf{w}^*(x, EI^*) - \hat{w}(x) \quad (8)$$

ermittelt.

2) *Regularisierung*: Wird lediglich das Funktional (8) mit realen Messdaten für das Optimierungsverfahren verwendet, erhält man instabile Ergebnisse. Entsprechend wird das Funktional der Zielfunktion erweitert. Es wird eine Regularisierung angewendet um eine stabile Lösung zu erhalten. Dem Funktional (8) werden die zwei Terme

$$\mathbf{F}_{\alpha, 1 \times j}(x, EI^*, \alpha) = \sqrt{\alpha} \mathbf{E} \mathbf{I}_j^*(x) \quad (9)$$

$$\mathbf{F}_{\gamma, 1 \times (j-2)}(x, EI^*, \gamma) = \sqrt{\gamma} \mathbf{L} \mathbf{E} \mathbf{I}_j^{*\top}(x) \quad (10)$$

hinzugefügt. \mathbf{L} ist eine finite Differenzen Matrix zweiter Ordnung. Die Regularisierungsparameter α und γ müssen iterativ ermittelt werden und sind kritisch für die Stabilität des Optimierungsverfahrens. Mit dem Parameter $j = 100$ resultiert das Funktional der Zielfunktion (7) somit in

$$\mathbf{F}(x, EI^*, \alpha, \gamma)_{1 \times 298} = \begin{bmatrix} \mathbf{F}_{\mathbf{w}} & \mathbf{F}_{\alpha} & \mathbf{F}_{\gamma} \end{bmatrix}. \quad (11)$$

3) *Parameterermittlung*: Zur Ermittlung der Regularisierungsparameter wird das Optimierungsverfahren für eine unterschiedliche Anzahl an Stützstellen n_{opt} und unterschiedlichen Regularisierungsparametern α und γ durchgeführt. Um die Güte der Approximation zu beurteilen, wird der mittlere absolute prozentuale Fehler

$$MAPE_{EI} = \frac{100\%}{n} \sum^n \frac{|EI_{real} - EI^*|}{EI_{real}} \quad (12)$$

verwendet. Zur Bestimmung dieses Fehlers muss allerdings die reale Biegesteifigkeit bekannt sein. Diese Voraussetzung ist bei einem Alpinski nicht gegeben. Entsprechend werden die Parameter basierend auf Messobjekten mit bekannter Biegesteifigkeit gewählt. Die Parameterermittlung wird in Abschnitt III-B umgesetzt.

III. ERGEBNISSE

Die Ergebnisse, welche mit den Methoden aus Kapitel II ermittelt wurden, werden in diesem Kapitel beschrieben. In Abschnitt III-A wird das Verfahren der Biegesteifigkeitsbestimmung anhand eines Flachstahls mit gleichmäßiger Geometrie detailliert beschrieben. Die Regularisierungsparameter, welche benötigt werden, um die Steifigkeit von Alpinski zu ermitteln, werden anschließend in Abschnitt III-B bestimmt. Dazu werden sechs Flachstähle mit variabler Breite analysiert. Abschließend wird die Biegesteifigkeit eines Alpiskis mit den erhaltenen Regularisierungsparametern bestimmt.

Die verwendeten Messobjekte aus Stahl wurden mit einem Laser aus einem 6 mm Stahlblech ausgeschnitten. Die Messobjekte mit variabler Biegesteifigkeit sind in Abbildung 3 dargestellt. Um die Biegesteifigkeiten EI_{real} zu erhalten, wurden die Geometrien der Flachstähle in einem Intervall von 100 mm nachgemessen. Mit dem Flächenträgheitsmoment I_y und einem E-Modul von $E = 2.1 \times 10^5 \text{ N/mm}^2$ ergibt sich der Verlauf der Biegesteifigkeit EI_{real} .

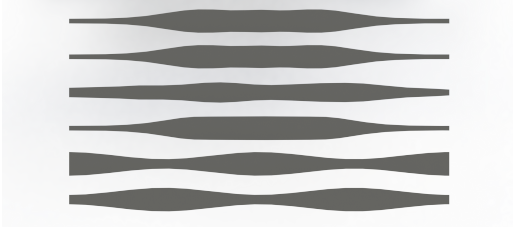


Abbildung 3. Messobjekte aus Stahl mit veränderlicher Breite respektive veränderlicher Biegesteifigkeit. Diese wurden mit einem Laser aus einem 6 mm Stahlblech mit einer Länge von $L = 1600$ mm ausgeschnitten. Die Messobjekte werden zur Ermittlung der Regularisierungsparameter verwendet.

A. Messobjekt mit konstanter Biegesteifigkeit

Die Bestimmung der Biegesteifigkeit aus realen Messdaten wird in diesem Abschnitt mit einem Flachstahl mit konstanter Geometrie durchgeführt. Die Abmessungen betragen $L = 1800$ mm, $b = 100$ mm und $h = 6$ mm. Die auftretende Kraft beträgt $\hat{F} = 105$ N.

1) *Wahl der Regularisierungsparameter:* Die Regularisierungsparameter werden, wie in II-B3 beschrieben, ermittelt. Die Werte für den mittleren absoluten prozentualen Fehler (12) wird mit den Parametern

$$\begin{aligned} n_{opt} &= \{10, 20, 30, 40\} \\ \alpha &= \{0, 10^{-30}, 10^{-29}, \dots, 10^{-14}\} \\ \gamma &= \{0, 10^{-20}, 10^{-19}, \dots, 10^{-10}\} \end{aligned} \quad (13)$$

untersucht. Wie viele Stützstellen notwendig sind, um die Biegesteifigkeit abzubilden, wird mit der Konvergenz von $MAPE_{EI}$ in Abhängigkeit von n_{opt} untersucht. Dies wird mit Abbildung 4 umgesetzt. Daraus ist ersichtlich, dass ab einer Anzahl von 20 Stützstellen keine signifikante Verbesserung der Approximation der Biegesteifigkeit mehr erreicht wird.

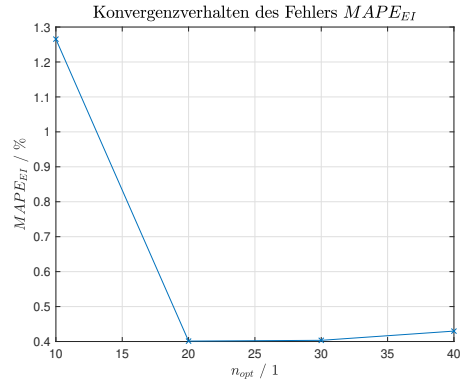


Abbildung 4. Messobjekt Stahl mit konstanter Biegesteifigkeit - Darstellung des mittleren absoluten prozentualen Fehlers $MAPE_{EI}$ der realen und angepassten Biegesteifigkeit für unterschiedliche Anzahl an Stützstellen n_{opt} und optimalen Regularisierungsparametern.

Für $n_{opt} = 20$ wurden mit den Regularisierungsparameter $\alpha = 1 \times 10^{-19}$ und $\gamma = 1 \times 10^{-10}$ die beste Approximation der Biegesteifigkeit erreicht.

2) *Ermittlung der Biegesteifigkeit:* Mit den ermittelten Regularisierungsparameter und der festgelegten Anzahl von Stützstellen wird die Biegesteifigkeit ermittelt. In Abbildung 5 ist die ermittelte Biegesteifigkeit an den Stützstellen und die interpolierte Funktion $EI^*(x)$ dargestellt. Der verbleibende Fehler der Biegesteifigkeit beträgt

$$MAPE_{EI} = 0.41\%. \quad (14)$$

B. Parameterermittlung mittels Messobjekten mit veränderlicher Biegesteifigkeit

Um ein stabiles Verfahren zu erhalten ohne die Regularisierungsparameter jedes mal zu ermitteln, werden die Messobjekte aus Stahl mit variierender Breite aus Abbildung 3 analysiert. Die analytisch berechneten Biegesteifigkeiten der Messobjekte sind in Abbildung 6 dargestellt. Mit diesen Biegesteifigkeiten EI_{real} wird die approximierte Biegesteifigkeit EI^* verglichen.

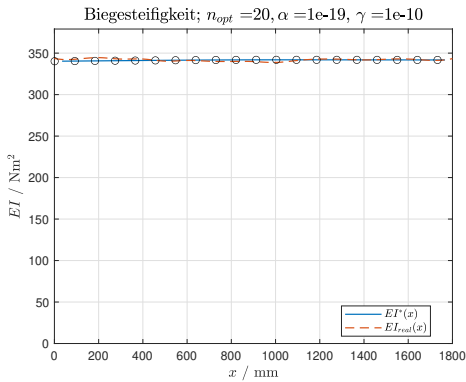


Abbildung 5. Messobjekt Stahl mit konstanter Biegesteifigkeit - ermittelte Biegesteifigkeit EI^* . Die Werte an den Stützpunkten sind in Schwarz markiert. Zum Vergleich die analytisch berechnete Biegesteifigkeit EI_{real} .

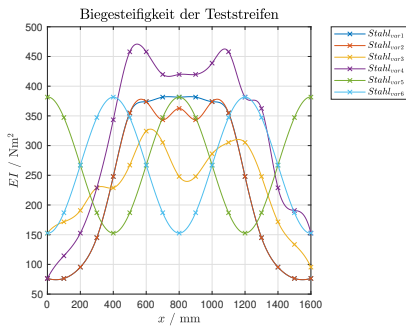


Abbildung 6. Analytisch ermittelte Biegesteifigkeit EI_{real} der Messobjekte aus Stahl mit veränderlicher Breite. Die Messobjekte sind in Abbildung 3 dargestellt. Die rekonstruierten Biegesteifigkeiten EI^* werden mit den Biegesteifigkeiten EI_{real} verglichen, um die Regularisierungsparameter zu ermitteln.

1) Wahl der Regularisierungsparameter: Die Messobjekte werden ebenfalls mit den Regularisierungsparametern (13) untersucht. Um die Anzahl der zu verwendenden Stützstellen n_{opt} zu bestimmen, wird die Konvergenz der Fehler $MAPE_{EI}$ in Abbildung 7 untersucht. Aus der Abbildung ist

ersichtlich, dass sich die Approximation beinahe aller Biegesteifigkeiten mit der Anzahl der Stützstellen verbessert. Der Fehler konvergiert beispielsweise bei der Messreihe $Stahl_{var4}$ nur leicht.

Da die Ergebnisse aus Abbildung 7 kein eindeutiges

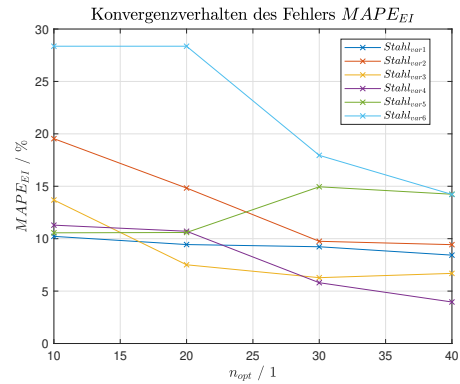


Abbildung 7. Messobjekte aus Stahl mit veränderlicher Biegesteifigkeit - Darstellung des mittleren absoluten prozentualen Fehlers $MAPE_{EI}$ für unterschiedliche Anzahl an Stützstellen n_{opt} und optimalen Regularisierungsparametern.

Konvergenzverhalten aufweisen, wird die Biegesteifigkeit im Weiteren an $n_{opt} = 40$ Stellen approximiert.

Als nächstes werden die Regularisierungsparameter analysiert. Ziel ist es, die Parameter α und γ so zu wählen, dass für alle Messobjekte aus Abbildung 3 gute Ergebnisse erzielt werden. Dies wird mit einem Punkte-Bewertungsverfahren umgesetzt. Der Wert mit dem geringsten Fehler $MAPE_{EI}$ erhält die höchste Punktzahl. Pro Datenreihe werden die 100 besten Ergebnisse mit dem Punktesystem bewertet. Die Punkte der besten 100 Werte werden über alle sechs Messreihen $Stahl_{var1}$ bis $Stahl_{var6}$ aufsummiert. Mit den Regularisierungsparametern $\alpha = 1 \times 10^{-28}$ und $\gamma = 1 \times 10^{-15}$ wurden die meisten Punkte erreicht.

C. Ermittlung der Biegesteifigkeit eines Alpinski

Abschließend wird die Biegesteifigkeit von zwei Alpinski analysiert. Von den zwei Slalomski ist lediglich bekannt, dass sie eine unterschiedliche Biegesteifigkeit aufweisen. Bei der Auswertung werden die ermittelten Parameter $n_{opt} = 40$, $\alpha = 1 \times 10^{-28}$ und $\gamma = 1 \times 10^{-15}$ verwendet.

Die Ergebnisse der ermittelten Biegesteifigkeiten sind in Abbildung 8 dargestellt. Der Ski wurde mit der Spitze bei $x = L$ gemessen. An den Randstellen wird die Biegesteifigkeit gegen 0 angenähert. Deshalb werden die Ergebnisse $EI^*(x < (x_A + 100 \text{ mm}))$ und $EI^*(x > (x_B - 100 \text{ mm}))$ in Abbildung 8 nicht dargestellt.

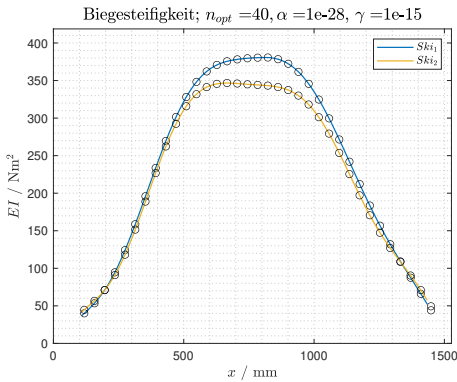


Abbildung 8. Messobjekte Alpinski - Vergleich der ermittelten Biegesteifigkeiten EI^* der beiden Alpinski. Für den Vergleich wurde die jeweils beste Approximation (bewertet nach $MAPE_w$) ausgewählt. Die ersten und letzten 100 mm der ermittelten Biegesteifigkeit sind nicht im Diagramm dargestellt.

IV. DISKUSSION

Die Qualität der Ermittlung der Biegesteifigkeit hängt von der Messgenauigkeit der Biegelinie ab. Die mittleren absoluten Fehler der gemessenen und approximierten Biegelinie

$$MAE_w = \frac{1}{n} \sum \frac{|\hat{w} - w^*|}{\hat{w}} \quad (15)$$

aller angeführten realen Messungen aus Kapitel III sind in Tabelle I angeführt. Im Vergleich dazu werden die Werte der Messgenauigkeit des Prüfstandes aus Abschnitt II-A3 betrachtet. Die Performance des Optimierungsverfahrens ist mit der Qualität der Messdaten ausgeschöpft. Eine bessere Approximation der Biegesteifigkeit könnte mit einer höheren Messgenauigkeit des Prüfstandes erreicht werden.

Messobjekt	MAE_{1w}	MAE_{2w}	MAE_{3w}
EI_{const}	22.69 μm		
$Stahl_{var1}$	12.89 μm	14.85 μm	13.17 μm
$Stahl_{var2}$	18.94 μm	15.07 μm	14.95 μm
$Stahl_{var3}$	16.50 μm	12.64 μm	17.18 μm
$Stahl_{var4}$	12.96 μm	(28.00 μm)	15.19 μm
Ski_1	14.14 μm	11.12 μm	12.59 μm
Ski_2	15.89 μm	14.51 μm	16.01 μm

Tabelle I
ABSOLUTER VERBLEIBENDER FEHLER MAE_w DER APPROXIMIERTEN UND GEMESSENEN BIEGELINIEN ALLER REALEN MESSUNGEN AUS KAPITEL III.

Zudem wird mit dem Funktional (8) der Fehler der Biegelinie lediglich an 100 Werten ermittelt. Dadurch ist das Verfahren anfälliger gegen „Ausreißer“ in den Messergebnissen.

Für das Messobjekt mit konstanter Biegesteifigkeit konnte eine ausgesprochen gute Approximation der Biegesteifigkeit erreicht werden. Mit den Messobjekten mit einer variablen Biegesteifigkeit konnten die Parameter der Regularisierung bestimmt werden. Mit diesen Regularisierungsparametern konnte für die Messreihen $Stahl_{var1}$ bis $Stahl_{var4}$ ein durchschnittlicher Fehler $\overline{MAPE}_{EI} = 9.81\%$ erreicht werden.

Durch die unbekannt reale Biegesteifigkeit der Alpinski sind die Ergebnisse aus Abschnitt III-C nur schwer zu interpretieren. Die Güte der Approximation kann lediglich über den Fehler der Biegelinien abgeschätzt werden. Vergleicht man die Fehler MAE_w aus Tabelle I ist eine ähnliche Performance der Approximationen sichtbar. Wie erwartet ist eine Differenz zwischen den Biegesteifigkeiten aus Abbil-

dung 8 erkennbar.

V. CONCLUSIO

Mit der Variante der iterativen nichtlinearen Regularisierung konnte ein stabiles Approximationsverfahren der Biegesteifigkeit entwickelt werden. An den Randbereichen konnte die Biegesteifigkeit allerdings nur schwer ermittelt werden. Zudem war die Ermittlung von lokalen, kleineren Änderungen der Biegesteifigkeit mit dem vorliegenden Prüfstand und der entwickelten Auswertung nicht möglich. Um die Biegesteifigkeit an den Randbereichen zuverlässig zu ermitteln, könnte eine höhere Messgenauigkeit des Prüfstandes Abhilfe schaffen.

Ein anderer Ansatz wäre, das Optimierungsverfahren zu verbessern. Mit einer variationalen Formulierung des Optimierungsalgorithmus respektive der Regularisierung, kann das Optimierungsproblem in ein Differentialgleichungssystem übergeführt werden. Durch den variationalen Formalismus können die Regularisierungsparameter mit einem Optimierungsverfahren bestimmt werden.

Auch wenn durch die Optimierungen die Biegesteifigkeit mit den vorgeschlagenen Maßnahmen noch genauer ermittelt werden könnte, verbleibt die besondere Herausforderung der Ermittlung der Torsionssteifigkeit eines Alpinskis.

DANKSAGUNG

Ein besonderer Dank gebührt Franz-Josef Falkner, der diese Arbeit betreut und mich stets unterstützt und motiviert hat. Ebenfalls möchte ich mich bei Robert Eberle für die Unterstützung bei der Implementierung der Regularisierung bedanken. Außerdem möchte ich Hannes Schwendinger für die Fertigung und Entwicklung des Prüfstandes.

Abschließend möchte ich mich bei meiner Familie und meiner Lebenspartnerin Marion Weiskopf bedanken. Ohne sie wäre es mir nicht möglich gewesen diese Arbeit zu verfassen.

LITERATUR

- [1] H. A. Schwendinger, "Experimental determination of the bending and torsional stiffness of an alpine ski," Master's thesis, Management Center Innsbruck, 2021.
- [2] D. Gross, W. Hauger, J. Schröder, and W. A. Wall, *Technische Mechanik 2 - Elastostatik*. Berlin Heidelberg New York: Springer-Verlag, 2017.
- [3] S. I. Kabanikhin, *Inverse and Ill-posed Problems - Theory and Applications*. Berlin: Walter de Gruyter, 2011.
- [4] J. Truong, C. Brousseau, and A. L. Desbiens, "A method for measuring the bending and torsional stiffness distributions of alpine skis," *Procedia Engineering*, vol. 147, pp. 394–400, 2016, the Engineering of SPORT 11. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1877705816307731>
- [5] F. Wikerman, "Characterisation of alpine skis," Master's thesis, KTH, Lightweight Structures, 2016.
- [6] T. T. Marinov and A. S. Vatsala, "Inverse problem for coefficient identification in the euler-bernoulli equation," *Computers & Mathematics with Applications*, vol. 56, no. 2, pp. 400–410, 2008. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0898122108000254>
- [7] R. Eberle and M. Oberguggenberger, "A new method for estimating the bending stiffness curve of non-uniform euler-bernoulli beams using static deflection data," *Applied Mathematical Modelling*, vol. 105, pp. 514–533, 2022. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0307904X22000051>
- [8] J. L. Mueller and S. Siltanen, *Linear and Nonlinear Inverse Problems with Practical Applications*. Philadelphia: SIAM, 2012.



Florian Mikula studierte am MCI Innsbruck/Österreich im Studiengang Mechatronik. Bei der Firma ematic in Landeck/Österreich ist er als Projektleiter im Bereich Mechatronik / Automatisierungstechnik tätig.

Implementation of monocular depth cue in 3D Software by the means of motion parallax and eye-tracking

Florian Mütterlein, Sebastian Repetzki (supervisor)

Abstract—Motion parallax offers a none hardware or marker dependent access towards 3D vision in common software applications such as CAD software, graphics software or game engines. With the off-axis projection algorithm, the transformation of the virtual object's orientation and translation can be adjusted towards the position of the human. Furthermore, this type of 3D vision promises to be more easy accessible due to the lack of marker or special glasses. In this paper, the implementation of such an algorithm into a self developed 3D graphics software is presented. The user's position for the algorithm is captured via an ordinary webcam. To test the software, measurements in terms of accuracy, speed and user feeling were performed. The results of the individual tests showed the necessity of the real-time capability of such a program and the resulting 3D feeling that scales with the response time and the accuracy of the program. Therefore, it was shown that an improvement of 3D software can be achieved with ordinary computer tools like a webcam.

Index Terms—Eye tracking, Image processing, Head coupled perspective, Graphics pipeline

I. INTRODUCTION

IN an ever growing market of 3D display applications, companies seem to be more and more interested in presenting and releasing self-contained products. Big technology players like Apple with the Vision Pro headset or Sony with their spatial reality display are revolutionizing the market for spatial vision. While many of these products often use simple

S. Repetzki is with the Department of Mechatronics, MCI, Innsbruck, Austria, e-mail: sebastian.repetzki@mci.edu.

ideas, the prices are to exorbitant to make them accessible to the average citizen. In addition, existing frameworks are rarely developed for commercial purposes to achieve a similar result. To create a 3D effect on a 2D screen, a depth perception effect is basically needed. The best known effect for this is binocular vision. While this is not possible on a regular screen without additional hardware such as special glasses, Cutting et al. investigated that the depth perception through motion parallax is almost as effective [1]. Johnny Lee is considered a pioneer in the use of the motion parallax effect in combination with head tracking. With a Nintendo Wii remote control, he was able to create a 3D effect on his television [2]. Kooima uses the same principle in his work on off-axis projection. He mathematically described an algorithm utilizing matrix transformations that can be implemented into OpenGL to achieve this exact 3D feeling [3]. This type of rendering is also called head coupled perspective and is used by Li et al. to build a framework that renders an image on the computer depending on the position of the user's head. For this purpose, an ordinary webcam and red-green 3D glasses are used. The glasses are utilized as markers for tracking and for 3D rendering [4]. Using the technique of background subtraction and color based filtering Szkudlarek et al. utilized the head coupled perspective as an additional input for computer games [5]. Employing the off-axis projection Buchanan et al. proposes an implementation for a view dependent rendering system for mainstream use. The tracking of the user is done by face tracking using a combination

of the Haar and Lucas-Kanade algorithms [6]. However, the computing power of modern computers is usually sufficient to integrate eye tracking in addition to face tracking. Bailer et al. proposed an algorithm for real-time eye tracking. In their work, they first create a region of interest in the image which is determined by the face tracking algorithm. Then the eye tracking is applied to this specific area to search for the user's eyes. Also narrowing down search areas inside this face region seem to lead to a faster yet stable eye detection. A prediction algorithm is used to compensate for the delay caused by the processing of the images by the camera and the program [7]. The present paper adopts this approach of eye tracking by further introducing a measured quantity called eye distance to conclude the depth of an user. Its used to implement the motion parallax effect into a self developed 3D software. The implementation follows the concept of a transformation pipeline. The criteria for a good 3D experience are identified as delay of the program and accuracy of the position detection. Tests for these criteria are performed to provide an outlook for future work with indications for improvement. The following section II lists the methods used to implement the objectives of the work. Subsequently, the results of the work, especially the latency and accuracy measurements, are presented in section III. In the last section IV, further steps are presented and the content of the paper is summarized once again.

II. METHODS

The main concept of the program is to utilize matrix multiplication to project a point from virtual space onto the screen in a defined pattern, resulting in realistic movement behavior. The key to achieve this effect is to synchronize the movement of the real user with the virtual camera. The screen serves as a window through which the user can observe the scene, with the perspective dynamically adjusting to the user's head position.

A. Transformation pipeline

This section sums up all the different transformations into one transformation pipeline. The coordinate transformation from the frame coordinate system into the screen coordinate system of the pipeline can be seen in Figure 1 and is further elaborated in the following. The transformation pipeline

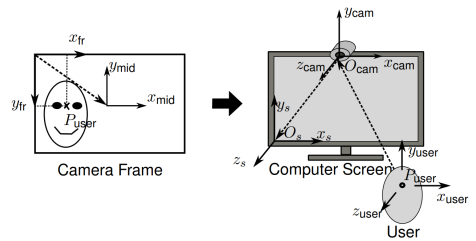


Fig. 1. Conversion from an input camera frame to a position ${}^s\vec{P}_{\text{user}}$ in screen space as first part of the transformation pipeline.

starts with the frame captured by the webcam, whose parameters are shown in Figure 2. The mid point between the two eyes ${}^{\text{fr}}\vec{P}_{\text{user}} = (x, y, 1)$ is determined by the means of an eye tracking algorithm. This vector lies in the frame coordinate system $\{0, x_{\text{fr}}, y_{\text{fr}}\}$ and is given in pixel coordinates. To transform the vector into the camera space $\{\vec{O}_{\text{cam}}, x_{\text{cam}}, y_{\text{cam}}, z_{\text{cam}}\}$ the coordinates are first transformed into the frame mid coordinate system $\{(\text{frame width}/2, \text{frame height}/2), x_{\text{mid}}, y_{\text{mid}}\}$.

$${}^{\text{mid}}\vec{P}_{\text{user}} = \begin{bmatrix} 1 & 0 & -\text{frame width}/2 \\ 0 & -1 & \text{frame height}/2 \\ 0 & 0 & 1 \end{bmatrix} {}^{\text{fr}}\vec{P}_{\text{user}} \quad (1)$$

By using the distance between the user's eyes the z coordinate of the user can be determined and therefore a mapping from the 2D frame mid space into the 3D camera space $\{\vec{O}_{\text{cam}}, x_{\text{cam}}, y_{\text{cam}}, z_{\text{cam}}\}$ can be performed. As can be seen in Figure 2, this paper refers to the distance between the user's eyes as the eye distance (ED) and the width of the user's face as the face width (FW). The following equations are

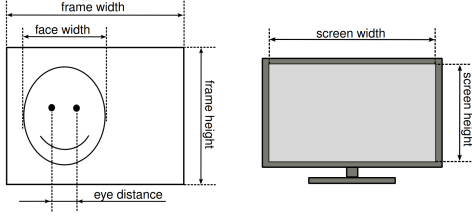


Fig. 2. Camera frame (left) and screen (right) parameters.

similar for both quantities and are thus discussed for the ED only. ED, measured in pixels, is inversely proportional to the user's distance from the camera and can therefore be mapped to the z coordinate of ${}^{\text{cam}}\vec{P}_{\text{user}}$ using a hyperbola $f_{\text{map}}(\text{ED})$.

$$f_{\text{map}}(\text{ED}) = a + \frac{b}{\text{ED} + c} \quad (2)$$

To simplify calibration, a hyperbola with just two coefficients is used. With two known Points ${}^{\text{cam}}\vec{P}_1$ and ${}^{\text{cam}}\vec{P}_2$ and their according ED the coefficients of the hyperbola can be derived.

$${}^{\text{cam}}\vec{P}_1 = (0, 0, z_1)^T \quad (3)$$

$${}^{\text{cam}}\vec{P}_2 = (0, 0, z_2)^T \quad (4)$$

$$z = \frac{b}{\text{ED}} - a \quad (5)$$

$$a = \frac{\text{ED}_2 \cdot z_2 - \text{ED}_1 \cdot z_1}{\text{ED}_1 - \text{ED}_2} \quad (6)$$

$$b = \frac{\text{ED}_1 \cdot \text{ED}_2 (z_2 - z_1)}{\text{ED}_1 - \text{ED}_2} \quad (7)$$

To transform the lateral position from the mid frame space to the camera space the ratio ${}^{\text{px}}r_{\text{cm}}$ between 1 pixel (short px) and 1 cm has to be determined. Because this ratio varies depending on the distance between the user and the camera, the ratio needs to be adjusted accordingly. By using the ED of the user measured in cm and the one measured in px the ratio can be calculated.

$${}^{\text{px}}r_{\text{cm}} = \frac{\text{ED} [\text{cm}]}{\text{ED} [\text{px}]} \quad (8)$$

Using the mapping function and the pixel to centimeter ratio the coordinates can be transformed into camera space.

$${}^{\text{cam}}\vec{P}_{\text{user}} = \begin{pmatrix} \text{mid } x_{\text{user}} \cdot {}^{\text{px}}r_{\text{cm}} \\ \text{mid } y_{\text{user}} \cdot {}^{\text{px}}r_{\text{cm}} \\ f_{\text{map}}(\text{ED}) \end{pmatrix} \quad (9)$$

As can be seen in the Figure 1 the coordinates of the user are needed to be in screen space $\{\vec{O}_s, x_s, y_s, z_s\}$. Note that the screen and camera orientation is aligned and their xy -plane is perpendicular to the negative z -axis of the user coordinate system. Using this the transformation from camera space into screen space is a translation along the origin vector ${}^{\text{cam}}\vec{O}_s$.

$${}^s\vec{P}_{\text{user}} = \begin{bmatrix} 1 & 0 & 0 & {}^sO_{\text{cam},x} \\ 0 & 1 & 0 & {}^sO_{\text{cam},y} \\ 0 & 0 & 1 & {}^sO_{\text{cam},z} \\ 0 & 0 & 0 & 1 \end{bmatrix} {}^{\text{cam}}\vec{P}_{\text{user}} \quad (10)$$

With the parameters screen width (sw) and screen height (sh) from Figure 2 the vectors v_a, v_b and v_c can be defined. They reach from the user towards the left upper (\vec{v}_b), left lower (\vec{v}_a) and right lower (\vec{v}_c) corner of the screen.

$$\vec{v}_a = -{}^s\vec{P}_{\text{user}} \quad (11)$$

$$\vec{v}_b = -({}^sP_{\text{user},x}, {}^sP_{\text{user},y} - \text{sh}, {}^sP_{\text{user},z})^T \quad (12)$$

$$\vec{v}_c = -({}^sP_{\text{user},x} - \text{sw}, {}^sP_{\text{user},y}, {}^sP_{\text{user},z})^T \quad (13)$$

With the vectors $v_r = (1, 0, 0)^T$, $v_u = (0, 1, 0)^T$ and $v_n = (0, 0, 1)^T$ the off-axis projection matrix according to Kooima can be calculated [3]. Figure 3 depicts the second part of the transformation pipeline. The virtual screen space is the virtual version of the real screen space and the virtual camera space is the virtual version of the user space. Having the position vector ${}^s\vec{P}_{\text{user}}$ the exact coordinates of the user towards the real screen are known. In the virtual world instead of moving the virtual camera, the objects of the virtual world are moved according to the movement of the real user and therefore the vector ${}^{\text{vcam}}\vec{O}_s$ from the virtual camera space $\{\vec{0}, x_{\text{vcam}}, y_{\text{vcam}}, z_{\text{vcam}}\}$ to the

virtual screen space $\{\vec{O}_{vs}, x_{vs}, y_{vs}, z_{vs}\}$ is equal to $-{}^s\vec{P}_{user}$.

$${}^{vcam}\vec{O}_{vs} = -{}^s\vec{P}_{user} \quad (14)$$

This translation of the objects is performed by

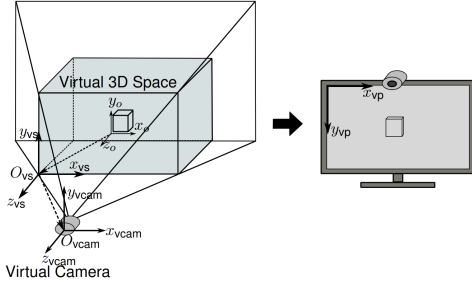


Fig. 3. Second part of the transformation pipeline to go from the user position ${}^s\vec{P}_{user}$ to a rendered pixel of a virtual object.

the off-axis projection matrix [3]. To be able to apply the projection onto the object coordinates, the objects position and orientation need to be formulated in virtual screen space. For this transformation an individual model matrix for every object needs to be formulated and multiplied onto the projection matrix. This multiplication is performed by a shader program.

B. Evaluation

1) *Latency*: The time the camera needs to send the image via USB to the PC and display it in the program is called hardware latency Δt_{hw} . In addition, the time that is needed to calculate the projection matrix and all other output values must be considered. It is mainly caused by the eye or face detection and is called Δt_{tr} . The time it takes the render program to perform all vertex transformations and to render the image accordingly is denoted by Δt_{rd} . Finally, the refresh rate of the screen may also cause the image not to be displayed immediately, but to wait for the refresh rate. This delay is formulated

as Δt_{fr} and denoted with its maximum value of 16.7 ms. The total program latency Δt_g is the sum of the previously introduced latencies.

$$\Delta t_g = \Delta t_{hw} + \Delta t_{tr} + \Delta t_{rd} + \Delta t_{fr} \quad (15)$$

To measure the hardware latency, the test setup shown in Figure 4 is used. On the laptop a program is running, referenced in the sketch as "Real frame", which opens a window that changes between black and white at any set frequency. Using a webcam, the window is filmed and displayed on the screen, in the picture referred to as "Captured frame". Furthermore, a timer displays the passed time in milliseconds. An external camera with at least 60 fps is used to film both frames. The time interval between the color change in the "Real frame" and the one in the "Captured frame" can be determined afterwards in the video. The transformation latency Δt_{tr} can be measured inside the software.

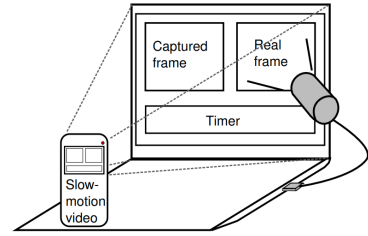


Fig. 4. Setup for the hardware latency measurement.

2) *Distance*: To measure the distance of the user towards the screen and therefore his z coordinate the setup from Figure 5 is used. The user's head is placed on a fixed base. The laptop with the webcam is then moved in 10 cm increments towards the user. The mapping function is created using FW or ED. The marked positions at every step are pos_{ref} while the values calculated via equation 5 are pos_{meas} .

3) *Trajectory*: For the trajectory measurement the setup from Figure 5 is used as well. The trajectory is only measured in z -direction and for that the

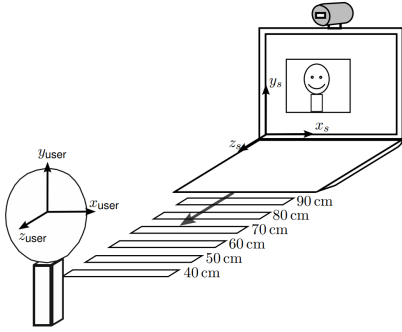


Fig. 5. Setup for the distance measurement.

computer is moved from a distance z_s towards another distance z_e at constant speed.

4) *Effectiveness*: To measure the effectiveness and the accuracy of the 3D effect the setup from Figure 6 is proposed. It can be seen that two real bars are connected to the laptop screen. The real bars are continued virtually in the program. In a realistic 3D experience, the edges of the virtual bars must match the running direction of the real bars and there must be no kink at the transition regardless of the user's position or movement.

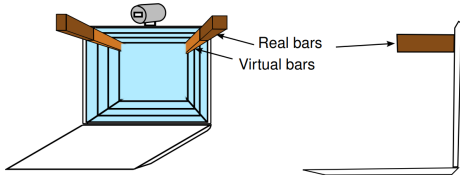


Fig. 6. Frontal (left) and side (right) view of effectiveness measurement setup.

III. RESULTS

A. Distance

In Figures 7 and 8 are the results of the distance measurements presented. In the diagram the values

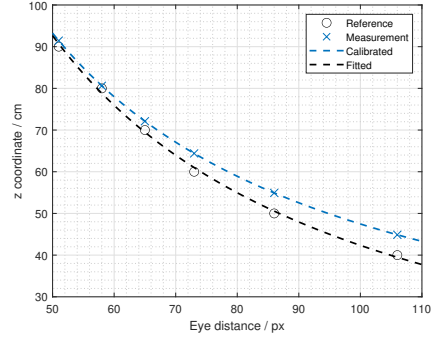


Fig. 7. Comparison of the calibrated and the fitted ED mapping function.

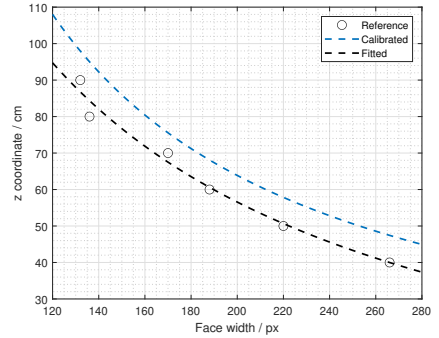


Fig. 8. Comparison of the calibrated and the fitted FW mapping function.

on the y -axis correspond to the distance between the user and the screen which is his z coordinate. The x -axis is the corresponding eye distance or face width in pixel measured at this point. The "o" points in the diagram form the reference values for the measured distances. Their z coordinate was measured with a tape measure. The ED or FW of these points corresponds to the mean value of 900 ED and FW measurements taken at this z coordinate. The "x" marks the median of values at a z coordinate that

are measured by to program with their respective ED. Because the FW mapping function just comes into play when the eye detection does not work no measurement points from the FW can be seen in Figure 8. All measurement distances are obtained by the ED mapping function. For this measurement the blue dashed mapping hyperbola was calibrated. The black dashed hyperbola is fitted through the reference points.

$$z_{ED,calibrated} = \frac{4577}{ED} + 1.2 \quad (16)$$

$$z_{FW,calibrated} = \frac{13243}{FW} - 2.85 \quad (17)$$

$$z_{ED,fitted} = -8.4365 + \frac{5101}{ED + 0.5} \quad (18)$$

$$z_{FW,fitted} = -20.89 + \frac{18777}{FW + 42.3946} \quad (19)$$

Figure 7 shows that the accuracy of the calibrated function increases for larger z coordinates. This can also be seen in Figures 9. They show the absolute error ϵ_{abs} between the calibrated mapping function and their fitted reference counter part. It can be seen that the absolute error for the ED mapping starts at almost 5.5 cm and is reduced to around 0.75 cm. Therefore the relative error of the z coordinate calculated by the ED mapping function drops from 13.5% to 1%. For the z -values calculated from the FW ϵ_{abs} rises from 7.5 cm to 12.5 cm. Its relative error ϵ_{rel} decreases from 18.5% to 13.5%. The measurements show that the ED mapping function is more reliable in terms of accuracy. Even though the accuracy at near distance is rather low with an error of up to 5.5 cm at a z coordinate of 40 cm the accuracy in the mid range is sufficient considering the choice of a low end webcam.

B. Trajectory

The results of the trajectory measurement can be seen in Figures 10 and 11. In the measurements the effects of a combination of post an pre-processing

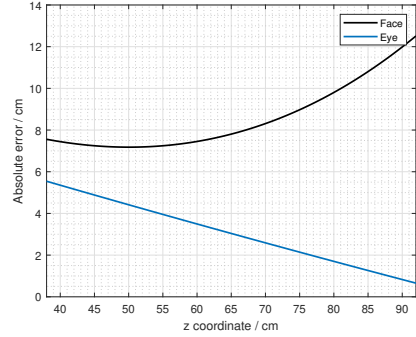


Fig. 9. Absolute error between the calibrated and fitted mapping functions.

were investigated. The data in Figure 10 is measured without post-processing and it can be seen that applying a Gaussian filter reduces the error between the reference trajectory and the measured value. As could be seen in the measurement before the noise is not reduced by this filter. The data in Figure 11 is post-processed with a moving average filter and pre-processed with a Gaussian filter with a kernel size of 5×5 . As expected from a moving average filter the noise reduction increases with the size of the filter. For slow trajectories no big delay can be observed.

C. Latency

The results of the latency measurement for the webcam are presented in Figure 12. The mean hardware latency $\Delta \bar{t}_{hw}$ from 25 measured quantities is calculated to be 101.83 ms. The standard deviation is 10.96 ms. Being a lower end webcam this result is expected. A possible improvement would be a custom exposure setting. Since not all webcam support this, the tests need to be done for the desired hardware. Figure 13 depicts the measurements of the transformation latency. The median $\Delta \bar{t}_{tr}$ is equal to 21.47 ms. The standard deviation is around 1.39 ms. The median of the render latency $\Delta \bar{t}_{rd}$ is measured to be around 0.5 ms. Because the value is so small

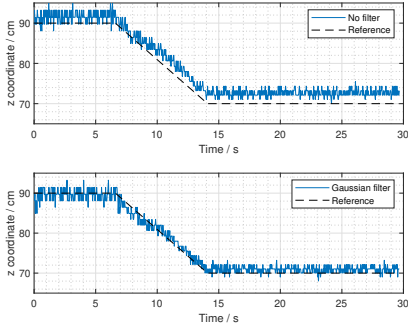


Fig. 10. Trajectory of the z coordinate with "No filter" (up) and "Gaussian filter" (down).

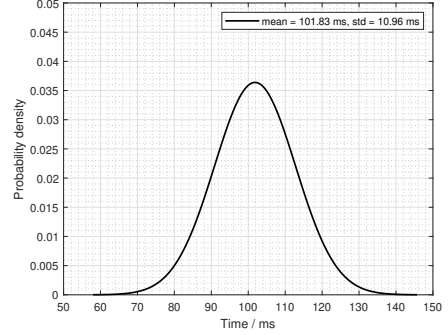


Fig. 12. Probability density function of the normal distribution for the hardware latency measurement.

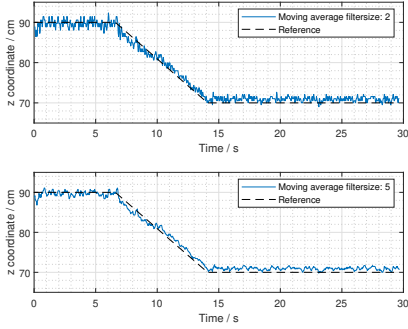


Fig. 11. Trajectory of the z coordinate with a moving average filter with size 2 (up) and size 5 (down).

in comparison to the other latency values or to their noise, it is neglected in the following calculations.

$$\Delta \bar{t}_g = (101.8 + 21.5 + 16.7) \text{ms} \approx 140 \text{ms} \quad (20)$$

In terms of reaction time the complete latency is noticeable and further improvement could be done by position prediction algorithms.

D. Effectiveness

The results of the effectiveness measurement can be seen in Figures 14. Since attaching the blocks onto

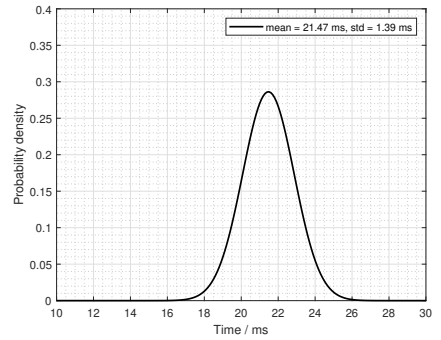


Fig. 13. Probability density function of the normal distribution for the transformation latency measurement.

the screen is rather challenging due to the limited possibilities for doing so, there is a small edge at the transition between the virtual and real blocks. Apart from this observation, the image shows that the position of the observer matches the one used for the projection and that the real blocks are virtually continued in the same direction. The transition of the edges from real to virtual block do not have any noticeable kinks.

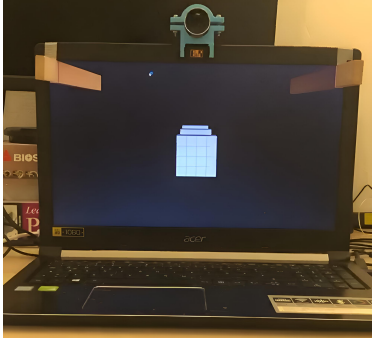


Fig. 14. Frontal view of the effectiveness measurement with wooden sticks attached to the screen.

IV. CONCLUSION

This paper proposed and evaluated algorithms to implement the motion parallax effect in 3D software. To track the position of the user in space relative to the computer screen the concept of the transformation pipeline was introduced. Further the accuracy and latency of the position detection was investigated. A calibration principle for a user specific mapping function was presented. The robustness of the mapping was increased by additionally use the face width for the mapping in case of a dysfunctional eye tracking. It was shown that for further distances from the screen around the origin of the xy plane of the webcam, the accuracy increases. Finally the projection was measured visually by means of a virtual extension of real objects. In the transition, there were no noticeable indications of an incorrect projection. For improvement of the latency measurements regarding different exposure values of the camera could be performed. Also a gpu acceleration for the eye tracking and face tracking functions could improve the speed and would need to be investigated further. To compensate for the negative effect of latency, position prediction algorithms could show promising results. Since the noise of the accuracy measurements originates in the user tracking, pre processing algorithms for noise reduction could improve

the 3D effect significantly. Also considering the user orientation towards the screen by some simplified assumption could already improve the accuracy further.

REFERENCES

- [1] J. Cutting and P. Vishton, "Perceiving layout and knowing distances: The interaction, relative potency, and contextual use of different information about depth," *Percept. Space Mot.*, vol. 5, pp. 69–177, 01 1995.
- [2] J. Lee, "Head Tracking for Desktop VR Displays using the Wii Remote." [Online]. Available: <http://johnnylee.net/projects/wii/>
- [3] R. Kooima, "Generalized perspective projection," 2011.
- [4] I. Li, E. Peek, B. Wünsche, and C. Lutteroth, "Enhancing 3d applications using stereoscopic 3d and motion parallax," vol. 126, 01 2012, pp. 59–68.
- [5] M. Szkudlarek and M. Pietruszka, "Head-coupled perspective in computer games," *Journal of applied Computer Science*, vol. 21, pp. 165–179, 2013.
- [6] P. Buchanan and R. Green, "Creating a view dependent rendering system for mainstream use." in *2008 23rd International Conference Image and Vision Computing New Zealand*, 2008, pp. 1–6.
- [7] C. Bailer, J. Henriques, N. Schmitz, and D. Stricker, "A simple real-time eye tracking and calibration approach for autostereoscopic 3d displays," *VISAPP 2015 - 10th International Conference on Computer Vision Theory and Applications; VISIGRAPP, Proceedings*, vol. 3, 03 2015.



Florian Mütterlein is a student of the "Mechatronics & Smart Technologies" program at the MCI Innsbruck/Austria. The topics of his work are largely related to computer vision and robotics. In his current master's thesis, he is also transferring this knowledge to the field of computer graphics.

Drone-Based 3D Mapping and Virtual Reality Inspection for Sport Facilities

Jan Prantner, BSc and FH-Prof. Bernhard Hollaus, PhD. (supervisor)

Abstract—This short paper explores the digitization of climbing walls using consumer-grade drones and Reality Capture photogrammetry software. The process involves capturing high-resolution images of the climbing wall and creating a 3D model that is rendered in Unity 3D engine. Additionally, a VR application is developed to allow for a virtual tour of the climbing wall. The digitization of climbing walls provides an immersive experience for climbers and enthusiasts and has potential applications in both recreational and competitive settings. The high-resolution 3D models can be used to develop customized training programs and custom holds for specific use cases, leading to more innovative and challenging routes. As technology continues to advance, there is great potential for further applications of digitization in the climbing world.

Index Terms—Climbing Walls, Digitization, Drones, Photogrammetry, Virtual Reality, Spatial Modeling

I. INTRODUCTION

THIS paper delves into the digitization of climbing walls, utilizing consumer-grade drones and photogrammetry software. The use of drones to capture high-resolution images and videos of the climbing wall and the subsequent creation of a 3D model provides an immersive experience for climbers and enthusiasts. The thesis goes further to explore the application of Unity as a 3D engine to render the 3D model in high resolution and the creation of a virtual reality (VR) application to allow for a virtual tour of the climbing wall.

The main objective of the thesis is to answer the research question: "How can consumer-grade drones

J. Prantner studies at MCI, e-mail: pj9625@mci4me.at.

and virtual reality technology be utilized to digitize and enhance the examination of climbing walls and other sports facilities?"

The final result should be an application which, in combination with virtual reality, enables interested users from the sports target group to examine a virtual replica of a sports climbing rock wall created from real data. On the basis of this, the foundation for more advanced applications, which are intended to offer added value in the areas of preparation, training and performance, is to be laid.

II. CONCEPTION

A. Methods for Digitising Real Objects

In the lead-up to this work, research was carried out on various methods for digitising real objects. The aim of this research was to find the most suitable approach for the application in this project, taking into account the available software and hardware resources.

In this context, the use of photogrammetry in combination with the software Reality Capture of the Slovakian company Capturing Reality [1] has turned out to be the most feasible based on the given circumstances.

B. Visualization Concept

Thus, the decision for the development environment of the virtual representation of the climbing wall falls on Unity 3D from the US developer Unity Technologies [2]. This choice is also backed by its cross-platform support, extensibility,

user-friendliness and the large community and resources available.

C. Final Concept & Time Schedule

Based on the selection made in the two subsections II-A and II-B regarding the software used for the photogrammetry and the visualization part as well as the technical means provided by the MCI, the resources used for the implementation of the project are summarised in the following table I:

TABLE I
KEY RESSOURCES & CORNERSTONES

Type	Description
Time frame	01.03.2023 - 31.07.2023
Photogrammetry software	Reality Capture 1.2
3D graphics software	Blender 3.4
Gamification engine	Unity 2021.3.25f1
Drone Model	DJI Mavic 3

The following table II shows a selection of the technical data of the with a focus on the on-board camera of the DJI Mavic 3, which is exclusively for all aerial shots.

TABLE II
CAMERA SPECIFICATIONS - MAVIC 3

Type	Description
Sensor	4/3 CMOS
Resolution	20 MP
ISO Range	100 - 6400
Aperture	$f/2,6 - f/11$
Focus Range	1 m - ∞
Shutter Speed Range	$\frac{1}{8000}$ s - 8 s

III. PHOTOGRAMMETRY WORKFLOW

A. Theoretical Background

Photogrammetry is a science and technology that involves obtaining reliable information about physical objects and the environment through the process of recording, measuring, and interpreting photographic images. It is primarily concerned with

making precise measurements of three-dimensional objects and terrain features from two-dimensional photographs.[3], [4].

- Taking photographs of an object from different locations and angles.
- Us software to develop "lines of sight" from each camera to points on the object.
- Mathematically intersecting these lines of sight to produce the 3-dimensional coordinates of the points of interest.
- Combining enough overlapping images of the same features to generate point clouds of topographic surfaces.
- Generating 3D digital meshes of the object.
- Development of the 3D model from the mesh including texture from the image material.

B. Data Acquisition

To implement the actual project, a suitable rock climbing wall had to be chosen. For practical reasons, the search was limited to the area directly surrounding Innsbruck. Criteria such as accessibility, approach time, bushes on the rock face and trees as well as other obstacles near the wall were taken into account when selecting the wall.

Against the background of the above-mentioned criteria, the Höttinger Steinbruch was chosen as the rock face. As it can be seen in Figure 1, the left part of the wall is particularly suitable for taking pictures for photogrammetry by drone due to its smooth rock and lack of vegetation.



Fig. 1. Höttinger Steinbruch climbing crack

In the course of the image acquisition, 265 close-up images were taken of the wall section that is to be reproduced in the greatest detail. The parameter combination of ISO 400 and an exposure time of $\frac{1}{240}$ s turned out to be optimal. The ISO value and the exposure time were set manually in manual mode and the f -number was determined automatically by the software for each image. The resulting images have an f -number in the range of $f/4$ to $f/7$. The pictures were taken in a resolution of 5280×3956 and saved in the RAW format *.dng* and the compressed format *.jpg*.

In addition to the photos taken, a total of 11 videos of the rock face were captured. The ISO value of all these shots has been fixed at ISO 100. The exposure time, depending on the distance to the wall, varies between $\frac{1}{100}$ s, $\frac{1}{160}$ s and $\frac{1}{240}$ s. The value for the f -number is constant at $f/2.8$ for all recordings.



Fig. 2. Perspective view with camera positions

Figure 2 shows the final result of the automatic image alignment process of the leftmost part of the wall, which requires the greatest level of detail and the most camera angles due to the density of the climbing routes and the nature of the rock.

The videos have been recorded in *.mp4* format with High Efficiency Video Coding (HEVC) H.265 encoding. For use in 3D modelling, every 15th frame from each of the 11 videos was extracted using a Matlab Frame Extractor tool which is developed by the MCI Medtech Department. By the use of the image extraction tool, a total of 1014 frames with a resolution of 5120×2704 pixels each were extracted.

C. Image Processing

Based on the Image data collected, the 3D model of the rock face is to be generated from the acquired image data set with the use of the photogrammetry software Reality Capture. Out of the original 1279 images, only 588 were used for generating the point cloud.

In the following, the mesh generation took 49 minutes and 43 seconds, resulting in a mesh with 50.3 million triangles (tris).

Due to the size of the mesh, it is not suitable in its unprocessed form for further use in software such as Blender or Unity and must therefore be reduced in its number of tris. By using the *Simplify Tool*, the simplification results in a new mesh consisting of 458.1 k tris. This corresponds to a reduction in tris of 99.08 % with a visually almost imperceptible difference and is therefore highly efficient in terms of data minimisation.

D. Model Processing

For the calculation of the textures, the master model with its 50.3 million tris, thus the highest level of detail, is used by Reality Capture. The

normal and height maps are used to reproject the rich geometric details of the master model texture onto the simplified level of detail 0 model by the use of the *Texture Reprojection Tool*. For the creation of this reprojected texture only 49 % of the original texture from master model are used and results in a texture resolution of $0.001676 \frac{m}{texel}$

As a final step in the modelling process, the simplified model and its textures were re-exported from Reality Capture as *.obj* and imported into Blender to refine the textures. In Blender, the UV map with the texture information is examined for areas with poor information content and these are corrected with various texture paint tools. The following Figure 3 depicts the model of the climbing wall that has been created.

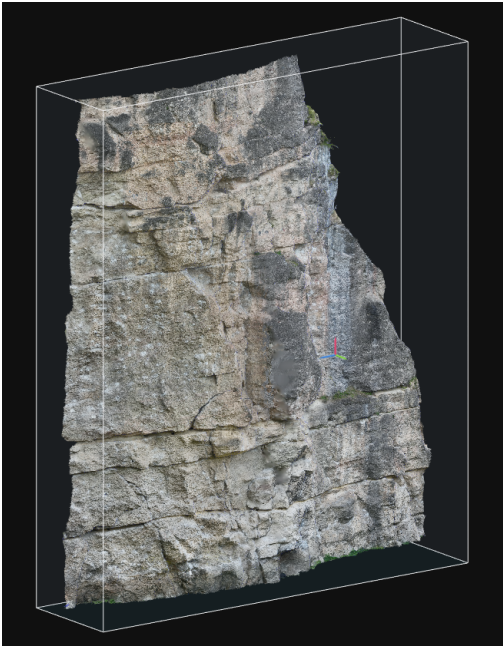


Fig. 3. Final climbing wall model isometric view

The following Figure 4 shows a section of the model of the rock face and demonstrates the level of detail of the final model achieved by applying the developed workflow.



Fig. 4. Climbing wall detail

IV. VISUALIZATION

A. Overview & Scope

The second major pillar of this project is the integration of the digitised rock face into an application that allows the user to examine it in an immersive way. The objective of this section is to create a virtual reality application that allows the model of the rock face created from the visual material to be examined.

V. UNITY PROJECT & WORLD

Unity High Definition Render Pipeline (HDRP) is used as the game engine and runtime for the realisation of the application.

As a base, a $256 \text{ m} \times 256 \text{ m}$ empty basemap was generated with the Unity extension Gaia 2021 [5].

In order to integrate the created model of the climbing wall into the terrain generated as described, a rock face of the mountain generated by Gaia must

be modified by using the *Terrain Tools*. The following Figure 5 shows the fully embedded climbing wall model in the constructed and designed game world. This virtual map serves as the basis for the VR user experience.



Fig. 5. Wall integration side view

A. Virtual Reality & Runtime

For the implementation of the virtual reality application, the *XR Interaction Toolkit (XRI)* version 2.3.2. provided by Unity is used.

In addition to the integration the classic VR input and output system consisting of a headset and two controllers, the *XR Device Simulator* is also implemented. This enables the virtual simulation of a VR head mounted display (HMD) output and the controller as long as it is activated. A separate control scheme is added to the standard XRI action map *XRI Default Input Action*. This reassigns the head tracking of the HMD to the mouse X and Y movement and the inputs and outputs of the controller are to different keyboard keys.

In this virtual environment, it is possible for the user to get an in-depth impression of this climbing wall and the climbing routes within it by using virtual reality hardware. The user is able to move freely in the virtual world and view the individual aspects and key points of the rock face in detail from the ground and flight perspective. Due to the moderate physical size of the crag and the game world as a whole, the hardware requirements

of the application could be kept within reasonable limits and allow operation on conventional consumer hardware.

The following Figure 6 shows a scene from the running game in the virtual reality mode of the application. The representation shown corresponds to the projection of the display rendered in the virtual reality glasses onto a classic screen.



Fig. 6. Full VR HMD projection

VI. SUMMARY & OUTLOOK

The premise and fundamental objective of this work was to answer the research question: "How can consumer-grade drones and virtual reality technology be utilized to digitize and enhance the examination of climbing walls and other sports facilities?".

In the course of the elaboration and answering of this question, a virtual world was constructed in which the climbing area of the Höttinger Steinbruch in Innsbruck is represented as a three-dimensional model that can be experienced by the user in a VR application. This example climbing area was used to show how the use of photogrammetry software is suitable for the virtualisation of real rock faces. In particular, whether the use of consumer grade drones and cameras makes it possible to capture image and video data of sufficient quality to create a model that is superior to classical route topos or

drawings of the rock face.

From a total of 1279 extracted frames taken from a series of 11 videos and 265 photographs, it was possible to generate a model of the rock face with a considerable degree of detail. In the course of the implementation, it was discovered that a well-planned combination of image acquisition parameters and the choice of suitable perspectives can drastically reduce the required amount of image material without negatively influencing the resulting level of model detail. In this case, the use of only 588 images of the acquired data set resulted in a sufficient quality of the final model. This finding positively validates the use of the workflow used here to generate practical and modern added value for the climbing sport in a relatively straightforward and efficient process.

By using Unity 3D to create a virtual environment for the inspection of the rock face, a VR application was also built which allows for a much more in-depth and realistic possibility of observation. This makes it possible for the user to examine certain places, such as key points in the climbing route, from different perspectives, as closely as possible as they would be found in the real world. Furthermore, it has been demonstrated that the operation of such an application is also possible on conventional hardware and also offers the potential, after sufficient optimisation, to be used on mobile devices, which further promotes its use as a tool for facilitating athletes.

The positive evaluation of the workflow used in this work opens up a number of further possibilities for the sport of climbing. A project directly linked to this work would be the creation of a digital register, similar to what already exists for classic climbing route topos, for virtual routes using VR for inspection. In addition, contributing users can mark important information about the individual routes separately, for example by selecting specific climbing

routes and noting detailed features such as key points and difficulty levels. The application can thus interactively provide guidance and highlight important features such as holds or footholds to help climbers plan their ascent.

ACKNOWLEDGMENT

I would like to express my gratitude to the supervisor of this work FH-Prof. Bernhard Hollaus, PhD for the straightforward and professional support in the realisation of this work. In addition, I would like to thank the staff member of the Medical, Health and Sports Engineering Department Jonas Kreiner, BSc. for the support with the hardware used.

REFERENCES

- [1] [Online]. Available: <https://www.capturingreality.com/>
- [2] [Online]. Available: <https://unity.com/>
- [3] K. Schindler and W. Förstner, "Photogrammetry," *Computer Vision*, p. 597–599, 2014.
- [4] K. Schwidefsky and F. Ackermann, "Photogrammetrie," 1976.
- [5] ProceduralWorlds. [Online]. Available: <https://www.procedural-worlds.com/products/indie/gaia/>



Jan Prantner is a Master student at the Management Center Innsbruck in the Mechatronics & Smart Technologies study program.

Optimizing the detection rate of machine learning algorithms for the detection of defects in prosthetic tooth production

Stefan Reinmüller and Bernhard Hollaus (supervisor)

Abstract—An important step in the prosthetic tooth production is the quality assurance. As this step is still done visually by employees, an automated quality assurance is needed. This procedure includes several tasks. The most important ones are taking high quality images with a homogeneous illumination, capturing the whole surface of a tooth and programming an algorithm that is able to detect the high variance of the occurring defects. The goal of this work was a holistic approach of optimizing these processes based on previous work.

In the first step, a prototype for taking images of the whole tooth surface was built. A polarizing filter and a dome light were used to get the best results. The second step was to train convolutional neural networks with images, made by the prototype. Additionally, artificial data was generated to improve the networks' accuracy.

The prototype was able to photograph the whole surface of a tooth without disturbances, like reflections or shadows. The trained neural networks reached accuracies of 96.4% for the defect classes under observation.

These results show that it is possible to automate the quality assurance of the defect class under observation in the prosthetic tooth production.

Index Terms—machine learning, CNNs, quality assurance, prosthetic tooth production, computer vision

I. INTRODUCTION

THE Austrian company Westcam Technologies GmbH is active as a supplier of automation solutions in the prosthetic tooth industry. An important step in the production process of dental prostheses is the visual quality control, which until now has been carried out mostly by hand. With classic image

S. Reinmüller studies at Management Center Innsbruck, Austria. e-mail: st.reinmueller@mci4me.at

processing methods, it is not possible to achieve a sufficient detection accuracy [1], [2]. Initial tests show that the use of machine learning could enable automated quality control for one occurring defect class [3]. Works like [4] and [5] show the capabilities of machine learning in the visual quality assurance of geometric variant objects like eggs and fruits.

The aim of this work is to optimize the detection rate of machine learning algorithms for the detection of optical defects in tooth production in two steps. To increase the amount of training data, the first step is to realize a prototype for automated recording of the tooth surface using one or more cameras. In the second step, the prototype will be used to record and label image data and then test various machine learning methods, in order to achieve accuracies above 95%, as required by the industry.

In the following sections, the needed image acquisition technology is defined. Based on this, the developed prototype, that is able to capture the whole surface of a tooth, is described. Finally, the results of data generation and the training of machine learning algorithms are presented.

II. METHODS

A. Image Acquisition Technology

Based on the image acquisition setups from [1] and [3] that lead to good results, an optimized setup was developed. In contrast to [3], the aim is to create a setup that enables automated image acquisition (including tooth handling). For this reason, the tooth

carrier is omitted. In addition, color defects must also be considered, so a color area scan camera is used.

Various experiments led to the following optimal test setup: The dome light used in [3] is provided with a linear polarizing filter as the light emission surface. In the center of the polarizing filter a hole is made to allow the light coming from the camera to pass through the dome light. The tube is optically opaque at the cylindrical surface of the dome light. At one end of the tube sits the camera together with its own linear polarizing filter (this must be aligned appropriately by rotating it to cancel the reflections on the tooth). The other end serves as the light inlet. The required components are listed in Table I. The schematic structure is shown in Figure 1.

TABLE I: Component list of optimized setup.

Pcs.	Description	Manufacturer
1	Camera: UI-1490SE-C-HQ	IDS
1	Lense: Lensagon CMFA2020ND, f = 20 mm	Lensation
1	Illumination: dome light DL097-WHI24	Advanced Illumination
1	Illumination Interior (light-tight tube)	self-made
1	Linear polarization film, 100 x 100 mm (XP42-18) (Art.Nr. 86179)	Edmund Optics
1	Lense polarization film: O-POL-27	Invisium

B. Prototype

The aim of the prototype is to automatically position a tooth for optimal exposure and photo capture from all sides. To find the best suitable concept for the prototype, different possibilities were compared. Thereby, a pairwise comparison, functional comparison and a Failure Mode and Effect Analysis are carried out. The best performing concept is the concept of the Single Tooth Glass Plate, shown in Figure 2.

Based on this concept, a final prototype is build. It is based on a profile frame made light-tight by

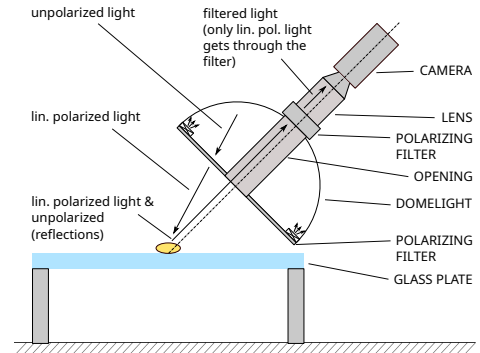


Fig. 1: Schematic representation of the optimized image acquisition setup. A tooth on a glass plate is captured by a camera at an angle of 45° . A dome light ensures the most homogeneous illumination possible. Reflections on the tooth are eliminated by using two linear polarizing filters. The (light) passage in the center of the dome light serves to shield the unpolarized light from the camera.

black wooden panels. Teeth are filled into a hopper through a closable opening in the lid. With the help of a vibratory feeder, the teeth are separated and are passed through the tooth feed chute at the center on the glass plate for photographing (or alternatively in the quick emptying box). Once the feed chute is folded back again, a photo is taken of 45° from above and from below. The glass plate is rotated by a desired angular amount and photos are taken again. This process is repeated until the entire tooth surface is photographed. For recording training and test data, the process ends here and the teeth are blown out into the reject box. For test operation of the algorithms, the blow-out can be done into an OK (label for a tooth / image without any defect) and a NOK (label for a tooth / image showing one or more defects) box. The glass plate including feed chute and optical components can be seen in Figure 4.

Resulting example pictures of the prototype are shown in Figure 3.

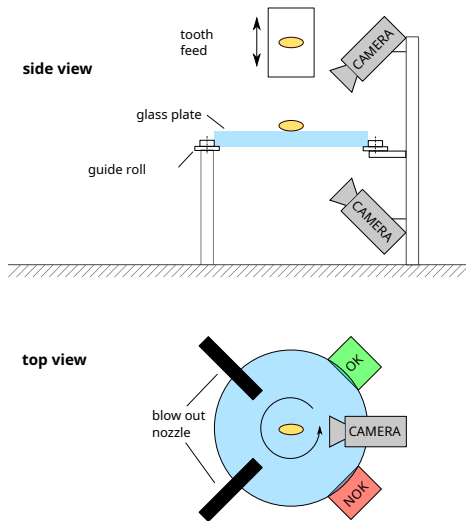


Fig. 2: Concept Single-tooth glass plate in side and top view: Teeth are centered on a rotating glass plate from above. The image is captured by two cameras, which are triggered after any rotation of the glass plate.

C. Image Acquisition & Labeling

Due to the best classifiability (defects can be detected very quickly by a human), the recording of the defect class dirty / dots, like depicted in Figure 3, is started. The following illumination and camera settings for maximum image quality are used:

- aperture setting upper camera: 16
- aperture setting lower camera: 22
- exposure time: 500 ms
- camera gain: 1
- lighting: 24 V continuous voltage.

Per tooth, 16 images per revolution are captured by each camera. This results in approximately 12000 images on which a defect of the defect class dirty / dots could be depicted.

Problems occur during the exposure due to wax residues on the teeth. This wax on the teeth comes

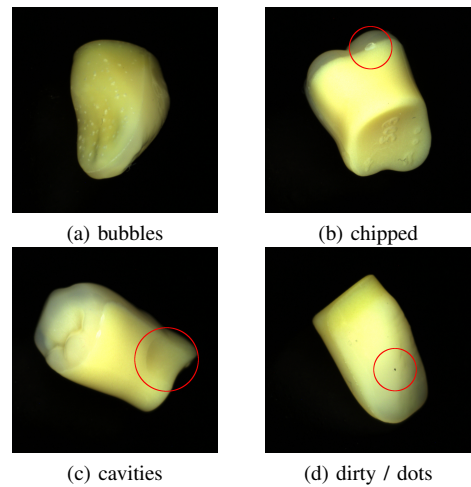


Fig. 3: (a) to (d): Example images of defect teeth captured with the optimized image acquisition setup.

from the current production process, since the teeth are put on wax strips before quality control. Despite repeated washing with hot soapy water (cleaning tip from the tooth manufacturer), an invisible wax film remains on the teeth, which are visually noticeable as defects. On the one hand, this allows dust and other particles to adhere to the teeth, which are visually noticeable as defects. On the other hand, each tooth releases some of the wax onto the glass plate of the abutment. After several hours of operation, a layer of wax and dust adheres to the glass plate, making almost all photos taken by the lower camera unusable. For a clean photo, the teeth must be cleaned, or corresponding non-waxed teeth must be used. A test run with about 8 hours of continuous use with wax-free teeth could confirm this. For this reason, the tooth manufacturer is trying to find a solution to this problem.

After sorting out unusable recordings, 6000 images of teeth of the class dirty / dots are left. Thereby, 1500 images show teeth with fine dust grains caused by the wax residue problem. Since the subsequently polluted teeth are differently judged by two different

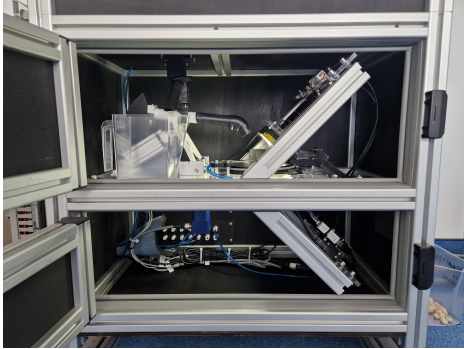


Fig. 4: With the front flaps open, the hinged feed chute (left in the picture) can be seen. In the position shown, the teeth are fed into the quick emptying box. If the chute is folded to the right by the cylinder, the teeth are positioned centrally on the glass pane. The camera setups are each located behind the profiles mounted at the 45° angle. In the lower opening, the drive motor can be seen, which drives the glass plate by means of a toothed belt.

people, these borderline unassignable examples are sorted out. Finally, there remain 2500 images with defect teeth and 2000 OK images. These images are split up in parts of 80 % for training and each 10 % for validation and testing.

D. Training Machine Learning Algorithms

Due to the high flexibility, as well as the current frequent use in research, the Python library PyTorch is used for the implementation of the algorithms. The following hardware / software setup is used for training and evaluation:

- Anaconda Version 4.14.0
- Development Environment: Visual Studio Code Version 1.79.1
- Python Version: 3.9.16
- torch Version: 2.0.1, torchvision Version: 0.15.2, CUDA Version: 11.8

To achieve high accuracies with a relatively small amount of training data, the state-of-the-art approach

of transfer learning is used. The idea of this concept is to use models that are pre-trained on millions of images, as a basis for further training on specific problems [6].

For image classification, processes are needed to handle the two-dimensional structure of an image efficiently. Thereby, it is important to obtain information, which results from pixels belonging together. The most common way of handling these problems are convolutional neural networks (CNNs) [6]. Based on this, the pre-trained convolutional neural network architectures ResNet [7], AlexNet [8], VGG [9], DenseNet [10] and EfficientNetV2 [11] are trained with the tooth dataset in this work. Further tests were made with adapting the input layers of the pre-trained networks to be able to process higher resolution images.

Additionally, to the classic CNNs, the architectures of vision transformers (ViTs) [12] and high resolution convolutional transformers (HCTs) [13] are tested.

An interesting tool to visualize the most important areas of an image, leading to the decision of a CNN, is gradient-weighted class activation mapping (Grad-CAM) introduced in [14]. For this procedure, the implementation from [15] is used.

E. Evaluating Machine Learning Algorithms

For evaluating the machine learning algorithms based on the test dataset, the confusion matrix is used. Thereby, the true positives (TP) are defined as samples which are correctly classified as OK and the true negatives (TN) are correctly classified NOK samples, whereas the false positives (FP) and false negatives (FN) represent misclassified samples. From the values of the confusion matrix, the true positive rate $TPR = \frac{TP}{TP+FN}$, the true negative rate $TNR = \frac{TN}{TN+FP}$ and the accuracy $acc = \frac{TP+TN}{TP+FN+TN+FP}$ are calculated. The values are used as percentages. For more clarity, the percentage symbol % is omitted in all tables.

III. RESULTS

A. Prototype Test

The testing of the prototype has revealed that the automatic capturing of images of teeth can be realized without any disturbances like shadows or reflections. Thereby, for people it is possible to safely detect coarse, medium and some of the very small flaws on the image alone (without seeing the real tooth).

Only for teeth with color defects, it is difficult to detect medium or slight appearances of this defect without any reference tooth. The reason for this are the different tooth lines, for some a "hard" color edge is okay and for others all edges must be "soft". When designing algorithms for this defect class in further research, approaches to reduce this problem, might be to separate the algorithm for each tooth line or color, or the additional information about which tooth line it is, is fed to a general algorithm.

B. Machine Learning Algorithms

1) Basic experiments with pre-trained networks:

To test the basic usability of pre-trained networks, some well-known networks are tested. Therefore, only the raw train data is used, without further extended methods.

All basic experiments lead to a training history like depicted in Figure 5. When training further, the models show the problem of overfitting (training accuracy stays at 100 % and the validation accuracy decreases). The best accuracy of 92.5 % is reached by resnet50, see Table II.

TABLE II: Evaluation measures of models tested in basic experiments with pre-trained networks.

Model	TPR	TNR	acc
resnet50	92.2	92.7	92.5
vgg11_bn	98.3	86.9	91.6
densenet121	96.7	83.8	89.1

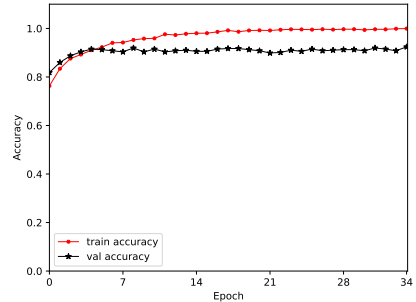


Fig. 5: Exemplary training and validation history of accuracy during training.

2) *Tests with data augmentation:* To reduce the impact of overfitting, when using small data sets, data augmentation is applied to retrain the best performing networks of the first test with geometry augmented images (scaled, rotated and moved).

The accuracy of two of the three trained models decreased in the training and test process (depicted in Table III). This could be the result of too few training epochs for the variance of the input data.

TABLE III: Evaluation measures of models tested with data augmentation.

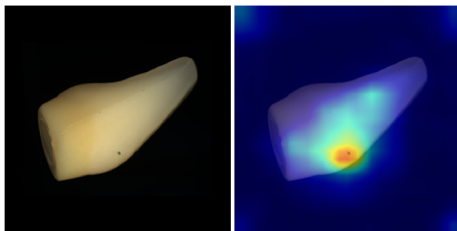
Model	TPR	TNR	acc
resnet50	94.4	90.3	92.0
vgg11_bn	96.7	86.5	90.7
densenet121	94.4	90.3	92.0

Additional tests using a learning rate scheduler in combination with an increase of the number of training epochs lead to a training accuracy of 100 %. The validation and test accuracy are still only slightly above 90 %. This indicates a limit to the generalization of the network on the basis of the number of training data, or a limitation by the pollution problem of the training data.

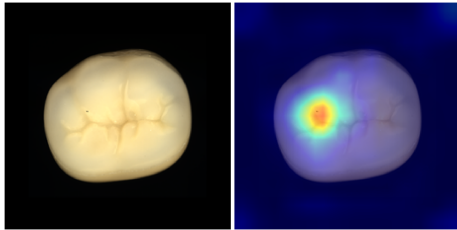
C. Grad-Cam Results

For a visual explanation of the network’s decisions, a grad-cam is applied, that shows the regions with the most impact to the classification result. Exemplary results of the trained resnet50 are depicted in Figure 6.

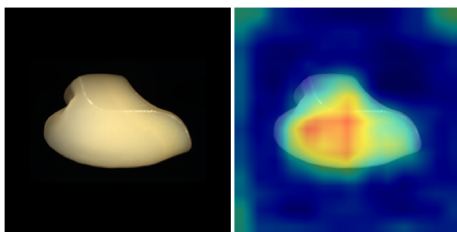
It can be seen that the network highlights the area around the defects as important. When passing an OK image, the grad-cam highlights almost the whole tooth surface. Small defects that are not detected also lead to a grad-cam image looking like an OK image.



(a) NOK - big errors



(b) NOK - medium errors



(c) OK image

Fig. 6: (a) to (c): On the left is the input image and on the right is the overlaid Grad Cam.

D. Experiments with higher input resolution

In order to better detect small errors, experiments are carried out in which the input dimension of the best performing network resnet50 is increased. Due to the structure of resnet50, the input dimension is theoretically variable. However, for a good performance in transfer learning, it is recommended to keep the processing dimensions of the convolutional layers.

To increase the input dimension of the network by a factor of two, the stride of the first layer of the original resnet50 is doubled (modified model: resnet50_highres_v1). By adding additional convolution layers, the input dimension of the network is increased by a factor of four (modified model: resnet50_highres_v2).

Furthermore, the small, medium and large version of the modern model architecture efficientnet_v2 with higher input dimensions were tested.

The higher input dimension leads to an accuracy increase, shown in Table IV. The best performing network was efficientnet_v2_s that reached an accuracy of 95.0%. The efficientnet_v2_l model reached the same accuracy, but with a higher false positive rate, making it unattractive for the real application.

TABLE IV: Evaluation measures of models tested with higher input resolution.

Model	TPR	TNR	acc
resnet50_highres_v1	52.2	73.4	64.7
resnet50_highres_v2	95.0	91.1	92.7
efficientnet_v2_s	97.8	93.1	95.0
efficientnet_v2_m	95.0	94.2	93.4
efficientnet_v2_l	95.6	94.6	95.0

E. Vision Transformer

Besides the classical CNNs, also other architectures become more popular. In this experiment the two pre-trained ViT models vit_l_14 and vit_l_16,

which are based on the architecture presented in [12], are tested.

Due to memory restrictions, not all parameters of the vit_l_14 model could be trained. This can be seen in the low accuracy compared to the accuracy of the vit_l_16 model, see Table V.

TABLE V: Evaluation measures of vision transformer models.

Model	TPR	TNR	acc
vit_l_14	82.8	84.9	83.8
vit_l_16	95.0	92.3	93.4

F. High Resolution Convolutional Transformer

As no improvement could be achieved with the ViTs, another architecture based on CNNs was tested. For experiments with the approach presented in [12], the code for the generation of the HCT model from the GitHub Repository to the paper was used. Since no pre-trained models are available, the models are trained from scratch.

The training accuracy of both models is under the 90% mark. The achieved accuracies for the validation and test data (see Table VI) are only about 70%. This indicates that with the available training data no sufficient generalization is possible. This would require a pre-trained model, or a correspondingly large amount of training data.

TABLE VI: Evaluation measures of high resolution convolutional transformer models.

Model	TPR	TNR	acc
htc_v1	65.6	74.1	70.6
htc_v2	67.2	70.7	69.3

G. Final Optimization

An attempt was made to further improve the efficientnet_v2_s network, which has performed best so far with an accuracy of 95.0%.

The best result is achieved by a combination of artificial noise, applied to the training images, and slightly adaptations to the training parameter. The test data shows that 5.4% of the class NOK is incorrectly identified as OK and 1.1% of the class OK is mistakenly sorted out as NOK. This results in an accuracy of 96.4%, as shown in Figure 7.

Accuracy: 96.4		predicted	
		OK	NOK
actually	OK	98.9	1.1
	NOK	5.4	94.6

Fig. 7: Relative confusion matrix of the final optimized efficientnet_v2_s model (scaled by the number of actual OK and NOK samples).

H. Overview of the achieved Accuracies

Figure 8 shows the best accuracies of all tested network architectures. The best accuracy of 96.4% was achieved with the efficientnet_v2_s model.

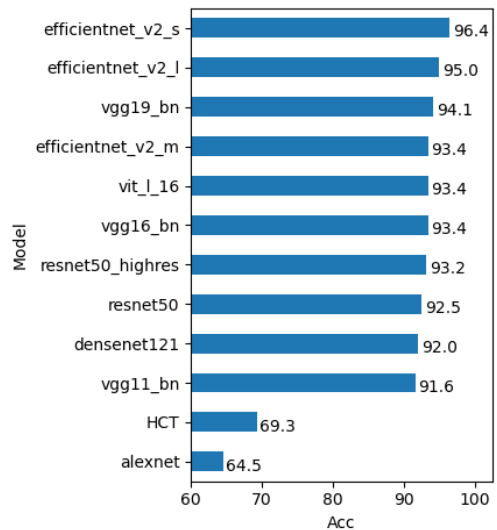


Fig. 8: Overview of the achieved Accuracies of the different models in percent.

IV. CONCLUSION

The best achieved results with an accuracy of 96.4% for the defect class dirty / dots shows the practical usability of machine learning for quality assurance in the prosthetic tooth production. Without the problem of wax residues on the teeth, even higher accuracies are imaginable.

For the complete automation of the visual quality control of denture teeth in the production plant, all occurring defect classes have to be integrated and the cycle time of the prototype has to be economically improved.

ACKNOWLEDGMENT

The author would like to thank the company Westcam Technologies GmbH, the involved prosthetic tooth manufacturer and the supervisor Bernhard Hollaus for enabling and supporting this project. Furthermore, huge thanks to Marleen Bahe for many discussions and the mental support during this work!

REFERENCES

- [1] Westcam Technologies GmbH, "Interne Arbeiten zum Thema Qualitätskontrolle in der Prothesenzahnherstellung," N.D., unpublished.
- [2] D. Schreter, "Concept of a Quality Inspection Process during Automated Manufacturing of Artificial Acrylic Teeth," Master's thesis, Management Center Innsbruck, Telfs, November 2013.
- [3] S. Reinmüller, "Qualitätssicherung in der Zahnproduktion durch den Einsatz von Machine Learning," Bachelor's Thesis, Management Center Innsbruck, Innsbruck, 2021.
- [4] A. Nasiri, M. Omid, and A. Taheri-Garavand, "An automatic sorting system for unwashed eggs using deep learning," *Journal of Food Engineering*, vol. 283, p. 110036, 2020.
- [5] A. Nasiri, A. Taheri-Garavand, and Y.-D. Zhang, "Image-based deep learning automated sorting of date fruit," *Postharvest Biology and Technology*, vol. 153, pp. 133–141, 2019.
- [6] J. Frochte, *Maschinelles Lernen: Grundlagen und Algorithmen in Python*. Carl Hanser Verlag GmbH Co KG, 2020.
- [7] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
- [8] A. Krizhevsky, "One weird trick for parallelizing convolutional neural networks," *arXiv preprint arXiv:1404.5997*, 2014.
- [9] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.
- [10] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4700–4708.
- [11] M. Tan and Q. Le, "Efficientnetv2: Smaller models and faster training," in *International conference on machine learning*. PMLR, 2021, pp. 10 096–10 106.
- [12] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly *et al.*, "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint arXiv:2010.11929*, 2020.
- [13] A. Taha, Y. N. Truong Vu, B. Mombourquette, T. P. Matthews, J. Su, and S. Singh, "Deep is a luxury we don't have," in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2022: 25th International Conference, Singapore, September 18–22, 2022, Proceedings, Part III*. Springer, 2022, pp. 25–35.
- [14] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-cam: Visual explanations from deep networks via gradient-based localization," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 618–626.
- [15] J. Gildenblat and contributors, "Pytorch library for cam methods," <https://github.com/jacobgil/pytorch-grad-cam>, 2021.



Stefan Reinmüller is a part-time student in the master program Mechatronics & Smart Technologies at MCI, Austria. Professionally, he works as a development engineer at the company Westcam Technologies GmbH (Austria) in the field of automation technology.

Methoden zur Durchführung Accelerated-Life-Time-Test am Beispiel eines Rasenmähers

Patrick Ruppert¹ and Franz-Josef Falkner² (Betreuer)

¹ Industrial Engineer A.M.I. GmbH and Management Center Innsbruck

² Lecturer Management Center Innsbruck Departement Mechatronik

Kurzfassung—Die Durchführung von Echtzeit-Ermüdungsschaden-Tests erfordern große personelle, zeitliche und dadurch finanzielle Ressourcen. Um Ressourcen einzusparen werden effektive Alternativen analysiert und evaluiert. Aktuell werden bei der Firma Stihl Rundlaufprüfanlagen verwendet, um die Ermüdungsschadennachweise zu erbringen. Diese Tests müssen aufgrund der Aussagekraft der Ergebnisse mehrfach durchgeführt werden, was dazu führt, dass enorme Ressourcen benötigt werden. Eine Alternative dazu sind computerunterstützte Software-Tools, welche eine Ermüdungslebensdauer-Vorhersage auf Basis von Messdaten im Zeit- oder Frequenzbereich erstellen. Mit Hilfe der verwendeten Methoden, welche hinter den Software-Tools stecken, kann eine schadensäquivalente Prüfung in einer deutlich verkürzten Dauer durchgeführt werden.

Bei der Evaluierung der beiden Domäne hat sich herausgestellt, dass sich beide Domäne für die Durchführung von beschleunigten Ermüdungsschaden-Tests verwenden lassen.

Die geeignetere Domäne ist der Frequenzbereich, da die Berechnung deutlich effizienter durchgeführt werden kann und die Reduktion der Testzeit größer ist als jene des Zeitbereichs. Bei der Bewertung der Zuverlässigkeit hat sich herausgestellt, dass auch die Variante im Frequenzbereich eine größere Zuverlässigkeit erzielt.

Schlagwörter—Ermüdungsschaden, Fatigue Editing, Fatigue Damage Spectrum, spektrale Leistungsdichte

I. EINLEITUNG

WERDEN Bauteile langfristig dynamisch beansprucht, kann es in Folge von stetig wachsenden Mikrorissen zum Bauteilversagen kommen.

Erste Mikrorisse können bereits durch Amplituden deutlich unter der Streckgenze entstehen. Damit Bauteile ihre geforderte Lebensdauer einhalten und somit eine hohe Zuverlässigkeit gewährleistet werden kann, werden Ermüdungsschaden-Tests durchgeführt.

A. Stand der Technik und Zielsetzung

Die Ermüdungsschaden-Tests werden aktuell in Echtzeit mit Hilfe von Rundlaufprüfanlagen durchgeführt. Die Rundlaufprüfanlagen bestehen aus einer kreisförmigen Grundplatte aus Beton, wo unterschiedliche Fahrbahnuntergründe montiert sind, die verschiedene Alltagssituationen eines Rasenmähers simulieren. Die Durchführung ist sehr zeitintensiv, wo durch enorme Ressourcen benötigt werden.

Als Alternative zu den bestehenden Methoden werden Methoden evaluiert, die es ermöglichen die Ermüdungslebensdauer zu berechnen. Auf Basis dieser Daten werden schadensäquivalente Profile erstellt, welche die Durchführung von Ermüdungsschaden-Tests in einer deutliche verkürzten Dauer ermöglichen. Durch die Reduktion der Dauer, können zeitliche, personelle und finanzielle Ressourcen gesenkt werden.

Um dieses Ziel zu erreichen werden die Methoden analysiert. Bei der Analyse wird der bereits etablierte Ansatz im Zeitbereich und der neuere Ansatz im Frequenzbereich untersucht. Bei der Evaluierung werden die Vor- und Nachteile bezüglich ihrer Effizi-

enz, Genauigkeit und Zuverlässigkeit der Ergebnisse bewertet.

B. Aufbau dieser Arbeit

In Abschnitt II werden die zwei Domäne Zeit- bzw. Frequenzbereich erläutert, mit deren Hilfe die Vorhersage der Ermüdungslebensdauer getroffen werden kann.

In Abschnitt III folgt die Bewertung der Methoden, wo die Ergebnisse und Erkenntnisse der vorgegangenen Abschnitte bewertet und erläutert werden.

In Abschnitt IV folgt die Schlussfolgerung, welches die geeignetere Domäne für die beschleunigte Ermüdungsschadenberechnung ist.

II. METHODEN

Um die Berechnung des Ermüdungsschadens zu beschleunigen, stehen zwei Domäne zur Verfügung, der Zeit- und der Frequenzbereich. Das Ziel der einzelnen Domäne ist die Reduktion der Prüfdauer unter den Rahmenbedingungen, dass das Ergebnis einen äquivalenten Schaden ergibt, wiederholt durchgeführt werden kann und die Belastungsumgebung wiedergegeben wird, sprich es zu keinem falschen Versagensmechanismus kommen kann.

A. Zeitbereich

Bei der Strukturanalyse von Zufallsschwingungen, wird zur Vorhersage der Ermüdungslebensdauer das gemessene Zeitsignal direkt analysiert. Grundlage der erfassten Daten bilden Dehnungen, welche aus Dehnmessstreifen stammen. Der *Zeitbereich* ist die einzige Domäne welche den Belastungsverlauf in Form von *Amplitude, Frequenz, Phasenbeziehung, Sequenz und Mittelwerteffekte* beinhaltet. Die Phasenbeziehung ist ein wichtiger Parameter, wenn es um die Ansteuerung von einem vier-Säulen-Prüfstand geht. [1]

Die Analyse ist sehr aufwändig, da zum einen der Informationsgehalt der Zeitsignale sehr groß ist und zum anderen werden Methoden verwendet, welche

numerisch berechnet werden. Aus diesen Gründen beschränkt sich die transiente Analyse auf kurze, diskrete Ereignisse. [2]

1) *Datenextrapolation*: Datenextrapolationsmethoden haben den Zweck die verkürzte Zeitdauer eines Messsignals zu extrapolieren. Eine Möglichkeit ist die sogenannte *Dauer Extrapolation*. Die Methode verwendet statistische Mittel und einen Kernel-Schätzer, um auf die Gesamtdauer zu schließen. Dazu wird das bestehende Rainflow-Histogramm in ein Wahrscheinlichkeitsdichtehistogramm umgewandelt. Abhängig von der Anzahl an aufkommenden Werte werden die Bandbreiten adaptiv angepasst, um anschließend das Signal zu extrapolieren. Mit Hilfe der Rainflow-Rekonstruktion kann das extrapolierte Zeitsignal hergestellt werden. Bei dieser Methode wird die Rainflow-Zählung rückwärts durchlaufen und Zyklus für Zyklus rekonstruiert. [3]

2) *Fatigue Editing*: *Fatigue Editing* ist eine Methode zur Verkürzung der Zeitdauer eines Ermüdungsschaden-Tests. Grundvoraussetzung für Zeitreduktionsmethoden ist, dass sich Ergebnisse auch bei wiederholter Durchführung reproduzieren lassen, sprich die Methode valide ist. Da die Eigenschaften bei dieser Methode erhalten bleiben, wird sichergestellt, dass das Ergebnis auch dem Anwendungsfall entspricht. Bei dieser Methode wird ein Zeitsignal in einzelne Abschnitte unterteilt. Mit Hilfe von *Omission von kleinen Zyklen*, werden Zyklen identifiziert und differenziert, ob ein Zyklus schädigend oder nicht-schädigend ist. In Abb. 1 ist die Anwendung abgebildet. [4]

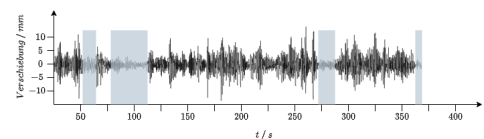


Abbildung 1. Identifizierung von nicht-schädigenden Anteilen [5]

Die nicht-schädigenden Anteile werden anschließend, wie in Abb. 2 dargestellt, aus dem Signal entfernt. [1]

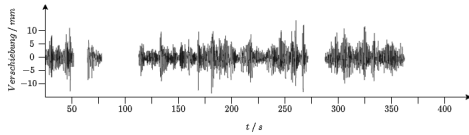


Abbildung 2. Darstellung von Unstetigkeitsstellen im Signal [5]

Da es keine theoretischen Modelle auf der Basis von Materialdaten gibt, muss das Entscheidungskriterium individuell festgelegt werden. Um zu verallgemeinern ab wann ein Zyklus als nicht-schädigend definiert wird, werden sogenannte *Filterstufen* basierend auf Ermüdungsfestigkeiten verwendet. Durch eine Vielzahl an Literaturen wird eine zulässige Filterstufe von 50 % der Dauerfestigkeit bestätigt. Somit werden alle Zyklen, welche diesen Grenzwert unterschreiten, als nicht-schädigend deklariert und aus dem Signal entfernt. Eine zusätzliche Reduktion der Zeitdauer kann erreicht werden, indem eine *Restschädigung* von beispielsweise 95 % vorgegeben wird. [1] [4] [5] [6]

Werden die nicht-schädigende Anteile aus dem Signal entfernt, entstehen im Signal *Unstetigkeitsstellen*. Um diese Unstetigkeitsstellen zu beseitigen, werden sogenannte *Verbindungsfunktionen* definiert. In Abb. 3 ist das kontinuierliche, zeitlich reduzierte Signal abgebildet. [1]

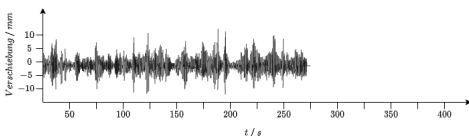


Abbildung 3. Kontinuierliches Signal nach einsetzen von Verbindungsfunktionen [5]

3) *Schädigungsrechnung*: Zur Berechnung der Ermüdungslebensdauer wird der dehnungsbasierte Ansatz auf Basis des Kerbgrundkonzeptes verwendet. Dabei besteht die Annahme, dass ein durch Ermüdung herbeigeführter Riss in einem Bauteil mit einem Ermüdungsbruch korreliert. Zu Beginn wird die Beziehung von *Coffin und Manson* verwendet, um die *Dehnungswöhlerkurve* zu beschreiben, welche den Zusammenhang zwischen der Gesamtdehnungsamplitude ε_a und den ertragbaren Anrisspielzahlen $2N_f$ liefert. Der mathematische Zusammenhang ist in Gl. (1) beschrieben. Die Gesamtdehnungsamplitude ε_a besteht aus einem elastischen Anteil ε_a^e und einem plastischen Anteil ε_a^p . [7] [8] [9]

$$\varepsilon_a = \varepsilon_a^e + \varepsilon_a^p = \frac{\sigma'_f}{E} (2N_f)^b + \epsilon'_f (2N_f)^c \quad (1)$$

Für die Berechnung werden der zyklische Schwingfestigkeitsexponent b , der zyklischen Duktilitätsexponent c , der zyklischen Duktilitätskoeffizient ϵ'_f , der zyklischen Schwingfestigkeitskoeffizienten σ'_f und das E-Modul E verwendet. [8] [9]

Für die Zählung der geschlossenen Hystereseschleifen wird die Rainflow-Zählung verwendet. Mit Zuhilfenahme der *Ramberg-Osgood-Beziehung*, des *Mansing-Verhaltens*, welche die Form eines Hysteresesastes nach der Belastung beschreibt und dem *Werkstoffgedächtnis*, welche die Fortsetzung des Pfades nach schließen einer Hystereseschleife beschreibt, können die Hysteresen bestimmt werden. Der Zusammenhang des zyklischen $\sigma - \varepsilon_a$ -Verhalten von Werkstoffen wird durch die *Ramberg-Osgood-Beziehung* wiedergegeben. Die Gesamtdehnungsamplitude setzt sich aus den Anteilen der elastischen ε^e und der plastischen Dehnung ε^p zusammen. Für die Berechnung der Gl. (2) werden die Spannungsamplitude σ_a , der E-Modul E , der zyklische Festigkeitskoeffizient K' und der zyklische Verfestigungsexponent n' verwendet. [8] [9]

$$\varepsilon_a = \varepsilon_a^e + \varepsilon_a^p = \frac{\sigma_a}{E} + \left(\frac{\sigma_a}{K'} \right)^{\frac{1}{n'}} \quad (2)$$

Um eine Beurteilung der Schädigung eines örtlichen Konzepts durchzuführen, wird die *Schädigungsparameterwöhlerkurve* aus der bestehenden *Dehnungswöhlerkurve* abgeleitet. Als Schädigungsparameter wird die Gesamtdehnungsamplitude ϵ_a herangezogen. Der Schädigungsparameter kann nach bekanntwerden der geschlossenen Hystereseschleifen, welche klassifiziert werden, mit den zugehörigen Spannungs- und Dehnungswerten an den jeweiligen Umkehrstellen berechnet werden. Der Schädigungsparameter P_{SWT} nach *Smith, Watson und Topper*, welcher sich nach Gl. (3) berechnet, ist von der Spannungsamplitude σ_a , der Dehnungsamplitude ϵ_a und der Mittelspannung σ_m abhängig. Das Produkt aus Spannung und Dehnung entspricht dem Äquivalent der Formänderungsenergiegedichte. [9] [10]

$$P_{SWT} = \sqrt{(\sigma_a + \sigma_m) \epsilon_a E} = \sqrt{\sigma_{max} \epsilon_a E} \quad (3)$$

Wird nun die Gl. (1) in die Gl. (3) eingesetzt und σ_a gleich null gesetzt, resultiert die *Schädigungsparameterwöhlerkurve*:

$$P_{SWT} = \sqrt{\sigma'_f (2N_f)^{2b} + E \sigma'_f \epsilon'_f (2N_f)^{b+c}} \quad (4)$$

Zum Schluss wird der Ermüdungsschaden, unter Modifikation der *kumulativen Schadenshypothese von Palmgren und Miner* und der Schädigungsparameterwöhlerlinie, berechnet. Ein Versagen eines Bauteils tritt ein, wenn die Summe der Schädigung gleich eins ist. [8] [9]

$$\sum_i D_i = \sum_i \frac{n_i(\epsilon_a)}{N_f(\epsilon_a)} \quad (5)$$

4) *Validierung*: Um zu überprüfen, ob die Methode valide ist, wird der Schaden vom originalen Zeitsignal und den mittels Fatigue Editing optimierten Signal miteinander verglichen. Kann ein Schadensäquivalenz nachgewiesen werden, so gilt dies als valide und das optimierte Signal kann verwendet werden, um den beschleunigten Test durchzuführen. [11]

B. Frequenzbereich

Mit Hilfe des *Frequenzbereichs* kann eine Vorhersage der Ermüdungslebensdauer getroffen werden. Bei diesem Verfahren wird das Zeitsignal nicht wie im *Zeitbereich* direkt analysiert, sondern mit Hilfe von *Wahrscheinlichkeitsdichtefunktionen* (englisch: Probability Density Function (PDF)), welche das Äquivalent zur Rainflow-Zählung im *Zeitbereich* ist, mit *Spektr*en und *spektralen Momenten* ausgedrückt. Eine effektive Methode um komplizierte Belastungen, wie sie in der Realität oft vorkommen, darzustellen, ist in Form einer spektralen Leistungsdichte. Dadurch wird die Rechenarbeit deutlich reduziert. Da die dynamischen Eigenschaften des Systems nicht bekannt sind, wird das System als Einmassenschwinger angenommen. Die Systemantwort wird aus dem Produkt der Eingangsgröße mit dem Quadrat der Transfer-Funktion berechnet. Somit wird durch das Miteinbeziehen der Systemreaktion das Antwortspektrum berechnet, ohne Kenntnis über den zeitlichen Verlauf der Anregung zu haben. Als Voraussetzung für die Verwendung des Frequenzbereichs muss die Zufallsschwingung normalverteilt, stationär und ergodisch sein. [2] [12] [13] [14] [15]

Strukturversagen resultiert aus übermäßiger Dehnungsenergie, welche bei schwingenden Bauteilen proportional zur Verschiebung ist. Auf dieser Basis folgt die Annahme, dass das Strukturversagen durch die größte Spannung im System herbeigeführt wird und äquivalent zur realen Spannung ist. Die Spannung σ im System ist proportional zur Relativverschiebung y . Der Proportionalitätsfaktor ist die Steifigkeit k . Dieser Zusammenhang ist in Gl. (6) beschrieben. [13] [14] [16]

$$\sigma = k y \quad (6)$$

Als Eingangsgröße werden Beschleunigungen verwendet, die aus Beschleunigungssensoren stammen, welche an repräsentativen Stellen am Rasenmäher angebracht sind. Für die Berechnung

der Ermüdungslebensdauer werden nur die Amplitude und Frequenz verwendet, da alle andere Informationen bei der Transformation in den Frequenzbereich verloren gehen.

1) *Transferfunktion*: Um die dynamische Antwort berechnen zu können wird angenommen, dass das Bauteil oder das System in Form eines linearen *Einmassenschwingers* (englisch: single degree of freedom (SDOF)) reagiert. Diese Annahme entspricht dabei nicht der Realität, da Systeme im seltensten Fall einem SDOF entsprechen. Um diese Annahme zu rechtfertigen, wird davon ausgegangen, dass die Reaktion von einem System von einem *einzigsten dominanten Modus* bestimmt wird, weshalb das System auf einen SDOF reduziert wird. Durch die Vereinfachung der Struktur auf einen SDOF wird die Struktur von seiner geometrischen Gestalt gelöst und wird somit standardisiert. Für die Bestimmung der Steifigkeit k und die Dämpfung d müssen die Systemsteifigkeit bzw. -dämpfung verwendet werden. Dies erfordert viel Erfahrung um die Systemwerte auf einen einzigen Parameter zu reduzieren. [12] [17]

Die Transferfunktion $H(f)$ wird verwendet um die dynamische Antwort eines linearen Systems abzuschätzen. Da die Transferfunktion eine komplexwertige Größe ist und nur der *Vergrößerungsfaktor*, sprich die Amplitudenüberhöhung von Bedeutung ist, wird der Absolutwert von der Transferfunktion $|H(f)|$ verwendet. Die Berechnung ist in Gl. (7) dargestellt. [12] [17]

$$|H(f)| = \sqrt{\frac{1}{\left(1 - \left(\frac{f}{f_n}\right)^2\right)^2 + \left(2\zeta \frac{f}{f_n}\right)^2}} \quad (7)$$

Dabei berechnet sich die dimensionslose *Lehr'sche Dämpfung*, auch *Dämpfungsverhältnis* genannt, ζ und die ungedämpfte Eigenfrequenz f_n mit der Einheit Hz wie folgt: [17]

$$\zeta = \frac{d}{2\sqrt{km}} \quad (8)$$

$$f_n = \frac{1}{2\pi} \sqrt{\frac{k}{m}} \quad (9)$$

2) *Shock Response Spectrum*: Das Shock Response Spectrum (SRS) repräsentiert die maximale Reaktion des Systems auf eine stoßartige Belastung. Transiente Stöße können z. B. durch ein Schlagloch, einen Zusammenstoß mit einem Objekt, das Be- oder Entladen des Mähers von einem Anhänger oder ähnliche stoßartige Belastungen verursacht werden. Die Methode bewertet die Schwere von Stoßbelastungen und wird für die Validierung der synthetisierten PSD verwendet. Für die Berechnung wird die Eingangsgröße mit der Transferfunktion mit gegebenem Dämpfungsverhältnis ζ gefiltert und das Maximum der Reaktion wird bestimmt. Dieser Vorgang wird durch das definierte Frequenzband wiederholt durchgeführt. Die Relativverschiebungen $y(t)$ werden für die Darstellung des SRS durch die *Pseudobeschleunigung* $\ddot{y}(t)$ ersetzt und gegen die Eigenfrequenzen f_n aufgetragen. Die Pseudobeschleunigung berechnet sich nach Gl. (10). [14] [18]

$$\ddot{y}(t) = (2\pi f_n)^2 y(t) \quad (10)$$

Das SRS wird für jeden Anwendungsfall separat erstellt. Um den Worst-Case zu repräsentieren und möglichst alle Fälle abzudecken, wird für die Lebensdauer-SRS eine Hüllkurve um alle SRS gebildet. [16]

In Abb. 4 ist das SRS in Form des *Relativverschiebung Schock Spektrum* $SRS(f_n)$ dargestellt. Im Diagramm wird die *Pseudo-Beschleunigung* mit der Einheit der Beschleunigung in m/s^2 gegen die Eigenfrequenz f_n mit der Einheit Hz aufgetragen. Die *Pseudo-Beschleunigung* repräsentiert nicht die Beschleunigung der Masse des SDOF, sondern ist nur eine Näherung. Die Beschleunigung der Masse stimmt nur mit der *Pseudo-Beschleunigung* überein, wenn die Dämpfung gleich null ist. Wird die Dämpfung jedoch als gering angenommen ($\zeta = 0.05$) darf die Näherung verwendet werden. Mit dieser Näherung entspricht die *Pseudo-Beschleunigung* gleich der Beschleunigung der Masse des SDOF. [18]

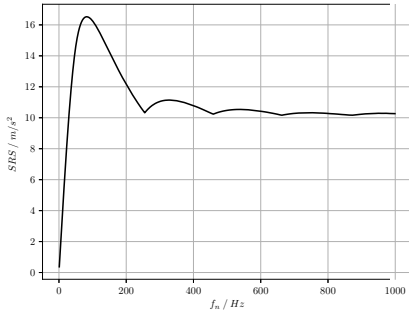


Abbildung 4. Relativverschiebung Schock Spectrum

3) *Extreme Response Spectrum*: Das *Extreme Response Spectrum* (ERS) repräsentiert die maximale Antwort von einem SDOF bei einem gegebenen Dämpfungsverhältnis ζ zu einem gegebenen Frequenzband, da die Eigenfrequenzen f_n des Systems nicht bekannt sind. Die Maxima, in Form von Relativverschiebungen, $y(t)$ repräsentieren die typischen Langzeit-Belastungen, welche ein Rasenmäher routinemäßig erfährt. Zu diesen gehören beispielsweise Bodenunebenheiten. Die Vorgehensweise ist im Prinzip identisch der des SRS. Somit unterscheidet sich das ERS nur durch das Eingangssignal vom SRS. Wie auch das SRS wird das ERS für die Validierung der synthetisierten PSD (englisch: Power Spectral Density (PSD)) verwendet. [13] [15]

Handelt es sich bei der Eingangsgröße um eine stationäre und normalverteilte Zufallsschwingung, kann eine Beschleunigungs-PSD $G_{\ddot{x}}$ als Eingangsgröße verwendet werden. Wird die PSD als konstant im Bereich der Resonanz angenommen, kann für die Berechnung der Reaktion in Form einer Relativverschiebungen y_{rms} , die Gl. (11) verwendet werden. Die *Pseudo-Beschleunigung* \ddot{y}_{rms} einer PSD berechnet sich aus dem Produkt der Relativverschiebung y_{rms} , welche in Gl. (11) beschrieben ist, und dem

Produkt aus $(2\pi f_n)^2$. [13]

$$\begin{aligned}
 y_{rms} &= \sqrt{\int_0^{\infty} |H(f)|^2 \frac{G_{\ddot{x}}(f)}{(2\pi f)^4} df} \\
 &= \sqrt{\frac{G_{\ddot{x}}}{64 \pi^3 f_n^3 \zeta}}
 \end{aligned} \tag{11}$$

Zu jedem Beschleunigungswert der PSD $G_{\ddot{x}_0}$ bei der Eigenfrequenz f_n wird die Relativverschiebung y_{rms} eines linearen SDOF berechnet. Dieser Vorgang wird für die Erstellung des $ERS(f_n)$ durch das komplette Frequenzband durchgeführt. Die Verwendung des *RMS-Wertes* für die Berechnung der Antwort ist nur innerhalb der Bandbreite der PSD gültig. Bei einer schmalbandigen Verteilung, welche eine *Rayleigh-Verteilung* der Amplituden entspricht, wird der Term in Gl.(11) um das Produkt $(2 \ln n_0^+ T)^{1/2}$ erweitert. Somit ergibt sich aus der ERS der Beschleunigung $ERS_{\ddot{y}}(f_n)$ folgender Ausdruck, welche auch als Gleichung von *Lalanne* bekannt ist: [13]

$$\begin{aligned}
 ERS_{\ddot{y}}(f_n) &= \sqrt{\int_0^{\infty} |H(f)|^2 G_{\ddot{x}}(f_n) 2 \ln(n_0^+ T) df} \\
 &= \sqrt{\frac{G_{\ddot{x}}(f_n) \pi f_n \ln(n_0^+ T)}{2 \zeta}}
 \end{aligned} \tag{12}$$

Wird die *Pseudo-Beschleunigung* gegen die Eigenfrequenzen aufgetragen, ergibt sich die in Abb. 5 gezeigte $ERS(f_n)$. Die Ordinate hat die Einheit der Beschleunigung in m/s^2 und die Abszisse die Einheit der Frequenz in Hz. [13]

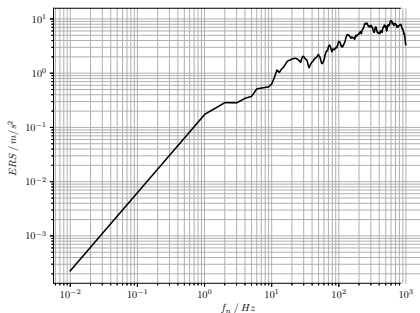


Abbildung 5. Darstellung eines Extreme Response Spectrums

4) *Fatigue Damage Spectrum*: Das *Fatigue Damage Spectrum* (*FDS*) repräsentiert das *schadensäquivalente Spektrum* zum Schaden, welcher durch die reale Anwendung entsteht. Bei der grafischen Darstellung des *FDS* werden die variierenden Schäden des Bauteils gegen die Eigenfrequenzen f_n dargestellt. Das Bauteil, bzw. die Struktur, wird durch einen linearen SDOF approximiert. Auch hier gilt der Hintergrund, dass die Eigenfrequenzen unbekannt sind und deshalb ein definiertes Frequenzband durchlaufen wird. Bei der Approximation des linearen SDOF wird ein festgelegtes Frequenzverhältnis ζ verwendet. Die Darstellung der Eingangsgröße in Form einer PSD der Beschleunigung $G_{\ddot{x}}(f)$ oder einem Zeitsignal der Beschleunigung $\ddot{x}(t)$ erfolgen. [13]

Ist die Eingangsgröße ein stationäre, ergodische, normalverteilte Zufallsschwingung mit Mittelwert null, kann mit Hilfe des analytischen Ansatzes das *FDS* berechnet werden. Die gemessenen Beschleunigungen werden mittels der *Welch-Methode* in eine PSD $G_{\ddot{x}}$ überführt. Die Systemreaktionen eines SDOF bei gegebenem Dämpfungsverhältnis ζ werden durch die RMS-Werte der Relativverschiebung y_{rms} dargestellt, welche in (11) beschrieben ist. [19]

[13]

Die Grundlage zur Berechnung des mittleren Schadens \bar{D} bildet die kumulative Schadenshypothese von Palmgren und Miner und die Basquin Gleichung. Hintergrund für die Verwendung des mittleren Schadens ist die Schwierigkeit der Bestimmung der Systemsteifigkeit k und der materialabhängige Parameter C der Basquin-Gleichung. Aus diesen Gründen werden die Werte gleich eins gesetzt. [13] [15]

$$\bar{D}_{NB} \approx \frac{k^b}{C} n_0^+ T (\sqrt{2} y_{rms})^b \Gamma\left(1 + \frac{b}{2}\right) \quad (13)$$

Für die Berechnung des *FDS*(f_n) wird der schmalbandige Ansatz verwendet. Um auf den Zusammenhang nochmals zu verdeutlichen wird zu jedem Beschleunigungswert der PSD $G_{\ddot{x}_0}$ bei der Eigenfrequenz f_n , der RMS-Wert der Relativverschiebung y_{rms} , welcher in Gl. (11) beschrieben ist, berechnet. Der berechnete Wert wird anschließend in die Gl. (13) eingesetzt und n_0^+ durch die Eigenfrequenz f_n ersetzt, um das *FDS*(f_n) zu berechnen. Der dynamische Vergrößerungsfaktor Q entspricht im Bereich der Resonanz $\frac{1}{2\zeta}$. [13] [15] [19]

$$FDS(f_n) = \frac{k^b}{C} f_n T \left(\frac{Q G_{\ddot{x}}(f_n)}{2(2\pi f_n)^3} \right)^{\frac{b}{2}} \Gamma\left(1 + \frac{b}{2}\right) \quad (14)$$

Eine grafische Darstellung des *FDS*(f_n) ist in Abb. 6 abgebildet. Die Ordinate hat die Einheit g^2/Hz und die Abszisse hat die Einheit Hz.

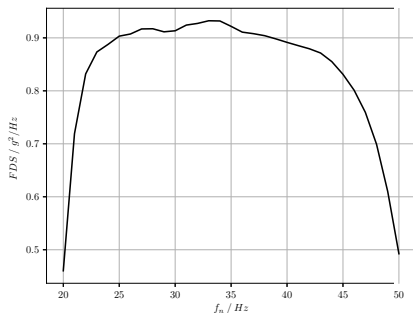


Abbildung 6. Darstellung des Fatigue Damage Spectrums

5) *Validierung im Frequenzbereich:* Um eine optimale Zeitreduktion zu erreichen, wird die synthetisierte PSD hoch skaliert, indem die Amplituden erhöht werden, bis das SRS den gleichen Schadensgehalt wie der gemessene Worst-Case-Fall enthält. Um zu gewährleisten, dass dies noch dem realen Anwendungsfall entspricht, wird eine Validierung durchgeführt. Bei der Validierung wird die synthetisierte PSD mit der Lebensdauer FDS des originalen Signals verglichen. Ein weiterer Schritt in der Validierung ist die ERS der synthetisierten PSD zu erstellen. Diese wird zum einen mit der SRS der Hüllkurve verglichen, um auszuschließen, dass eine zu große Skalierung stattgefunden hat. Zum anderen wird die ERS der synthetisierten PSD mit der Hüllkurve der ERS verglichen, um sicherzustellen, dass auch alle erfassten Maxima beachtet werden. [15]

III. BEWERTUNG DER METHODEN

1) *Zeitbereich:* Die Zeitsignale stammen aus Dehnmessstreifen, welche an repräsentativen Stellen befestigt werden. Die Auswertung der Daten ist sehr aufwändig. Durch den enormen Informationsgehalt des Signals sind auf der einen Seite alle Informationen enthalten, auf der anderen Seite kann nur eine kurze Zeitdauer aufgezeichnet werden. Die Methoden sind numerisch, weshalb die Methoden rechenintensiv sind. Die Rainflow-Zählung

erfolgt nicht sequentiell, da nur geschlossene Hystereseschleifen gezählt werden. Durch die Anwendung des Fatigue-Editing lässt sich eine realistische Reduktion der Zeitdauer auf Faktor zwei schließen.

Die Validierung ist ein Vergleich des Ermüdungsschadens. Bei dieser Methode gibt es keine Überprüfung, dass neben der Ober- auch die Untergrenzen des Bereichs sichergestellt wird, was zu einer Erhöhung der Zeitdauer führen kann.

2) *Frequenzbereich:* Die Messsignale stammen von Beschleunigungssensoren, welche am Rasenmäher angebracht werden. Durch die Transformation in den Frequenzbereich gehen Informationen, wie die Phasenbeziehung verloren. Der Frequenzbereich basiert auf den analytischen Methoden von PDF, PSD und spektralen Momenten, welche effiziente Methoden zur Verarbeitung von Zufallsschwingungen sind. Durch die Annahme des SDOF wird das System von seiner Geometrie gelöst, womit sich Systeme besser vergleichen lassen. Durch die Bildung von RMS-Werte lassen sich stoßartige Belastungen nicht original getreu abbilden.

Bei der Validierung mittels ERS und SRS wird sowohl die Unter- als auch die Obergrenzen der synthetisierten PSD überprüft, was zu einem zuverlässigen Ergebnis führt.

IV. SCHLUSSFOLGERUNG

Die in dieser Arbeit durchgeführte Evaluierung zeigt, dass sich sowohl der Zeit- als auch der Frequenzbereich für die Anwendung von beschleunigten Ermüdungsschadensberechnung verwenden lässt. Der vielversprechendere Ansatz ist der Frequenzbereich, da die Verarbeitung von Zufallsschwingungen in Form von PSD deutlich effizienter verarbeitet werden. Durch die Anwendung von PDF zur Beschreibung der Verteilung bzw. des Kollektivs können auch hochzyklische Belastungen sehr effizient bearbeitet werden. Beim Ansatz zur Bestimmung der beschleunigten PSD, welche aus den ermittelten Fahrdaten aus dem bestehenden Rundlaufprüfstand entnommen werden können, verspricht der Ansatz eine ideale Abbildung vom realen Anwendungsfall. Durch die

Validierung, bei der sowohl die Unter- als auch die Obergrenze abgedeckt werden, ist eine Zuverlässigkeit der Ergebnisse gegeben.

DANKSAGUNG

Der Autor möchte sich bei Thomas Kemmerich von der Firma nCode® und bei Holger Boller von Vibration Research beanken, die ihm bei einigen Themen hinsichtlich seiner Arbeit unterstützt haben. Der Autor dankt auch dem MCI für die Möglichkeit seine Masterarbeit dort zu schreiben.

LITERATUR

- [1] W. El-Ratal, M. Bennebach, X. Lin, and R. Plaskitt, "Fatigue life modelling and accelerated test for components under variable amplitude loads," in *Symposium on fatigue testing and analysis under variable amplitude loading conditions*, 2002.
- [2] R. Wang, *Frequency domain fatigue analysis of dynamically sensitive structures*. University of London, University College London (United Kingdom), 1997.
- [3] D. F. Socie and M. A. Pompetzki, *Modeling variability in service loading spectra*. ASTM International, 2004.
- [4] A. Halfpenny, "Methods for accelerating dynamic durability tests," in *Proceedings of the 9th International Conference on Recent Advances in Structural Dynamics, Southampton, UK*, 2006, pp. 17–19.
- [5] M. Mattetti, G. Molari, and A. Vertua, "New methodology for accelerating the four-post testing of tractors using wheel hub displacements," *Biosystems Engineering*, vol. 129, pp. 307–314, 2015.
- [6] P. Heuler and T. Seeger, "A criterion for omission of variable amplitude loading histories," *International Journal of Fatigue*, vol. 8, no. 4, pp. 225–230, 1986.
- [7] Y.-L. Lee, J. Pan, R. Hathaway, and M. Barkey, *Fatigue testing and analysis: theory and practice*. Butterworth-Heinemann, 2005, vol. 13.
- [8] J. Baldwin and J. Thacker, "A strain-based fatigue reliability analysis method," 1995.
- [9] M. Wächter, *Zur Ermittlung von zyklischen Werkstoffkennwerten und Schädigungsparameterwöhlerlinien*. Technische Universität Clausthal Clausthal-Zellerfeld, Germany, 2016.
- [10] B. Bohlmann, *Zur Schwingfestigkeit von Schweißverbindungen: das örtliche Konzept am Beispiel einer Rohrverzweigung*, 1995.
- [11] nCode, "User manual [software]," *Hottinger Bruel & Kjaer UK Ltd*, vol. 23.0.0 Release Build 403, 2023.
- [12] A. Halfpenny and F. Kihm, "Environmental accelerated testing. use of virtual test to extend physical approach," *White Paper, Astelab*, pp. 1–11, 2008.
- [13] C. Lalanne, *Mechanical vibration and shock analysis, specification development*. John Wiley & Sons, 2013, vol. 5.
- [14] M. A. Biot, "Transient oscillations in elastic systems," Ph.D. dissertation, California Institute of Technology, 1932.
- [15] A. Halfpenny, "Accelerated vibration testing based on fatigue damage spectra," *White paper, nCode International*, www.ncode.com, 2006.
- [16] A. Halfpenny and T. Walton, "New techniques for vibration qualification of vibrating equipment on aircraft," *Aircraft airworthiness & sustainment*, vol. 2010, p. 19, 2010.
- [17] J. S. Bendat and A. G. Piersol, *Random data: analysis and measurement procedures*. John Wiley & Sons, 2011.
- [18] C. Lalanne, *Mechanical Vibration and Shock Analysis, Mechanical Shock*. John Wiley & Sons, 2013, vol. 2.
- [19] —, *Fatigue Damage: Mechanical Vibration and Shock Analysis, Volume*. John Wiley & Sons, 2014, vol. 4.



Patrick Ruppert absolviert den Masterstudiengang Mechatronik und Smart Technologies am MCI Innsbruck/Österreich und arbeitet als Industrial Engineer in einem Medizintechnikunternehmen.

Developing a gas engine emission model for controls validation

Johan Cornelis Pieter Sanner, Georg Arnold (supervisor), and Benjamin Massow (supervisor)

Abstract—Emissions models for gas engines improve the usefulness of digital controller validation. Validating combustion engine controllers in a digital environment is becoming increasingly important. New regulations, as well as competition, are important for the development of new control strategies for internal combustion engines, with emissions being part of it. Emissions are influenced by many different properties of a combustion engine. Since gas engines are mainly used for energy generation, most of these engines operate with a lean burn concept to reduce emissions. Another benefit of gas engines is fuel flexibility, different gases can be used to operate the engine. The final emission model, must consider those influences. As basis for the model, available existing emission related controllers can be used. Combined with the emission models already existing and measuring data, a suitable model is derived. Gas mixing is not as widely discussed in those models, and will not be implemented in the final model. The created emissions model is then validated in digital environments which are already in use. Implementation in the mean value engine model (MVEM) represents the measurements quite accurately. After the MVEM, validation on a hardware in the loop (HiL) system is conducted. Based on the results of the MVEM and HiL, the emission model can be used for validation in future projects. Further investigation is needed for other engine series which are not accurately represented by the implemented model. The influence of gas mixing on emissions is not worked out since desired properties for a gas mixing model are not realizable with the given models.

Index Terms—mean value engine model, discrete engine model, hardware in the loop, combustion engine, nitrogen oxides, emission model, emissions, gas engine

I. INTRODUCTION

POWER grids are build up of different technologies with different dependencies. Solar energy needs the sun, wind energy will not work with an

absence of wind and other energy sources have other issues. Gas engines for energy generation can supply energy quickly, and support the grid. Gas engines can therefore bridge demands or fluctuations in power grids caused by the increased use of renewable energies. They can also supply electricity and heat to a community. By combining these energy sorts, a high efficiency can be achieved. With high efficiency lower running costs for the operator are possible.

To operate a gas engine, not only do the mechanical parts play an important role, also controls engineering is needed. Some parts that are needed for the operation of an engine depicted in Figure 1. Displayed are the air filter and gas dosing valve (1). At the connection of fresh air and gas a module to blend the gas with the air is located, which is then forced through the turbo charger. After which the mixture will be cooled in the intercooler. Controlling the speed, power and even emissions is done by the turbo bypass valve (2) and throttle valve (3). The turbo bypass valve will return some of the charged mixture back to the intake of the turbo charger. Power is created by a generator (4). Controlling heat output of an engine is done by a cooling water pump (5). By using the thermal energy released by an engine, other appliances can be served as well, and by doing so increase efficiency.

A lot of other components also come into play in operating a gas engine. This Figure only provides a small overview of the components present on a combustion engine. These components are used for governing power output, but also emissions. Other means of emission reduction are considered during the design phase of an engine, or done via exhaust

gas treatment.

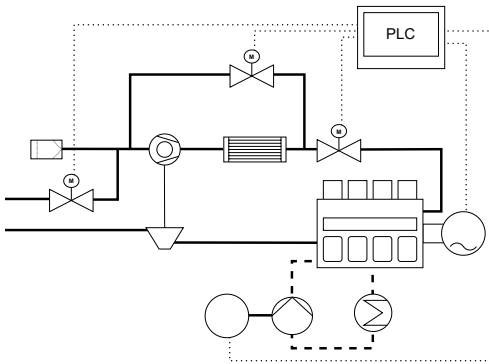


Fig. 1. Simplified schematic overview of a gas engine.

Since gas engines are operated all over the world, different regulations apply to them. To adhere to different regulations, new controller strategies must be developed. To implement new technologies, validation plays an important part. For an effective validation digital models are used. Which can save time and money spend.

Regarding emission controller strategies, an emission model for validation is needed. With the help of such a model, different controller implementations can be validated. This model will help adhering to regulations regarding emission standards. For gas engines the amount of nitrogen oxides (NO_x) in the exhaust gas is of interest. The final model needs to estimate the NO_x values present in the exhaust gas.

II. GAS ENGINE OVERVIEW

Gas engines can be adopted to different use cases. There are different gases that can be used to operate a gas engine, among others natural gas, all sorts of bio gases, mine gas and even hydrogen. Especially with some of the bio gases, there can be environmental and health issues because of their poisonous nature. By burning these gases, they are converted to different components, making them less harmful. To

burn those different kinds of gases, different engine properties are needed. Different pistons or turbos are only a couple of varieties that compose a gas engine. These alterations increase performance of an engine for a given gas.

Gas engines are mostly designed as lean burn engines. This means an excess amount of air is mixed to the fuel type in question. Because of the high level of air, some of the problematic gases that occur during combustion are already reduced. A graph depicting the emissions for a given air to fuel ratio can be seen in Figure 2. It can be seen, that the NO_x values are decreasing, with an increased λ value as opposed to carbon monoxide (CO) and hydro carbons (HC) emissions, which increase with leaner operation of the engine above the λ value of 1.6.

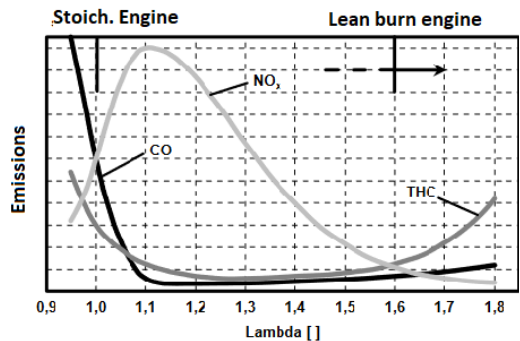


Fig. 2. Graph for emissions depending on the air to fuel ratio [1].

Through adaptation of different elements that make up an engine for different gases, other properties of the engine are also changed. Some of these properties have an impact on the emission values, some of which are listed below.

1) *Mixture creation and air fuel ratio*: Creation of the mixture for combustion can be achieved in different ways. There is premixed, intake manifold injection and alternatively direct injection.

For gas engines the premixed and intake manifold

injection are of importance. They allow for an excess amount of air in the combustion chamber. A lean air fuel ratio will reduce the amount of emissions. Gas engines are designed to operate with a λ value of over 1.6, where nitrogen oxides (NO_x) are decreasing, see Figure 2. The λ value represents the air to fuel ratio and can be calculated by rearranging

$$h_u = \frac{H_u}{1 + \lambda L_{min}} \quad (1)$$

where the mixture heat value, h_u is calculated by the heat value of the used gas H_u and the needed amount of air L_{min} .

2) *Compression ratio*: Another factor impacting emissions is compression ratio of the cylinder. This can be altered by adjusting the geometry of the piston. Decreasing clearance volume of a cylinder increases compression.

Creating higher a compression ratio results in higher combustion temperatures, which increase NO_x values [2]. This is where the benefits of a lean burn engine become apparent. Since there is more air available as necessary for the combustion, the excess amount of air reduces the temperature during combustion. To force more mixture into the combustion chamber, a turbo charger is used. Through the higher compression ratio, a higher air to fuel ratio is needed to keep emissions constant.

3) *Other influences*: There many other properties of an engine that change the emission and efficiency. By changing geometry of the combustion chamber, a different turbulence swirl into the cylinder can be achieved [2], with different filling and combustion patterns. Ignition timing is also worth considering. Changing the timing will influence the efficiency of the engine and in turn also the emissions [3].

A lower charge temperature increases performance of an engine. The mixture is cooled by an intercooler after its pressure is increased by a turbo charger. For different gases and engine geometries, the optimal temperature will vary. Different temperatures require different operation boost pressures to achieve the same emissions. But also it is used to achieve a better fill of the cylinder and reduce knocking.

Another technology is the exhaust gas recirculation (EGR). This returns some of the exhaust gases back into the intake manifold. The EGR is more commonly used in rich burn engines and is not needed for lean burn engines [4].

4) *After treatment*: If the amount of emissions created during combustion is not sufficiently low, emissions in exhaust gases are reduced by after treatment. One of the technologies are catalysts, e.g. selective catalytic reduction (SCR) catalysts. Different catalysts have different use cases. Other means consists of particle filters or even absorbers.

A. Controllers on a gas engine

To operate a gas engine, other than the mechanical components, controllers are also of importance. Governing speed of an engine in idle or during island operation is done by means of a speed controller. If an engine is connected to the grid the power controller takes over which includes keeping the engine in its rated operating window. But other properties of an engine also need controlling e.g. heat. To get rid of this heat, cooling must be present, which is also controlled. This is valid for cooling but also mixture cooling, which is kept at an optimal temperature for the engine to operate at.

Controlling emission values is important for the operation of a gas engine. A lot of different techniques can be used to reduce emissions. One way is by measuring different values in the intake and calculating the optimal operating point by doing so.

B. LEANOX controller

Keeping the values of the engine at a certain level can be done with the help of the LEANOX controller [5], with LEANOX being a trademark of INNIO group. The heat value of the mixture h_u can be calculated with

$$h_u = P_{mech} \frac{T}{p} \quad (2)$$

by measuring the values for pressure, p , and temperature, T , in the intake manifold, combined with

the actual power, P , of the engine. Together with the needed amount of air L_{min} for the heat value H_u of the gas from

$$h_u = \frac{H_u}{1 + \lambda L_{min}} \quad (3)$$

the needed air to fuel ratio can be calculated. For a stable combustion process this method works really well. With the calculated air to fuel ratio, the engine operates at a certain emission level, covering a lot of different influences. To setup the controller, some values are needed. Among these are the boost pressure, boost temperature, operating power and speed at half load and full load. By doing so, the optimal point for power output in between can be estimated and emissions stay at a certain level.

Since the properties of an engine change over its life time, a closed loop control is preferred, which is achieved by extending the LEANOX controller with an emissions sensor. By doing so, the changes in engine properties can be counteracted, stabilizing emissions.

C. Virtual validation

Implementing new controller strategies is needed to adhere to regulations and keep an edge over competition. Validation is part of development processes. To reduce costs and save time, large parts of validation can be achieved on digital models. Controllers are mostly implemented in MATLAB or even Simulink, which is also suited for digital validation. A lot of processes during combustion are represented by thermodynamic equations, which are used to create a digital engine representation. To reduce computational effort mean values are used, creating a MVEM. In a MVEM the processes of an engine are replicated, with which the controller is tested. Since the model of the engine, and the model of the controller both are executed in the same software and mostly on a personal computer (PC), there is no proper implementation in a real time environment. For this reason, validation on a hardware in the loop (HiL) system should also be

done before transitioning to a real engine. Which is a discrete implementation called discrete event model (DEM), where the crank angle is the independent variable [6].

On a HiL system the engine models are also represented but in a real time environment. The real time can be used to calculate the crank angle after every cycle. Since a HiL system only represents the engine, a programmable logic controller (PLC) is connected to it. Because of the differences in setup, the PLC runs a version of the control software which is close to the final software. This represents a real world engine more accurately as the MVEM.

D. Existing models

Since combustion engines are widely used, existing emission models are investigated. The Zeldovich mechanism [7] is used as a basis to derive emission models from. With the help of the Zeldovich mechanism the equation

$$\frac{d[NO_x]}{dt} = \frac{6e^{16}}{\sqrt{T}} \exp\left(\frac{-69090}{T}\right) [N_2]_e \sqrt{[O_2]_e} \quad (4)$$

is derived in [8] and the equation

$$\frac{d[NO]}{dt} = 2 k_{1f} \sqrt{36.64T} \sqrt{[O_2]} \exp\left(\frac{-27123}{T}\right) [N_2] \quad (5)$$

is derived in [9]. Both use the combustion temperature to estimate the nitrogen monoxides (NO) respectively NO_x values. Other implementations are also presented, but a simple implementation is targeted for the NO_x model that shall be implemented.

Since fuel composition also has an impact on emissions, a gas mixing model shall also be implemented. Some modelings for mixing of gases can be found, but those require a detailed composition of gas properties. This makes for a complex way of implementing a gas mixing model. For dual fuel engines with diesel and gas, a mixing model is required, but based on different fuels with different properties and a different engine geometry.

III. METHODS

A. Requirements

The emission model shall be easy to implement, work for different engine types and handle varying emission levels for different regulations in different countries. It is important that the behaviour of the emissions is replicated, not the precise values. One limitation, is that most available measurement data, which replicate different emissions, is only available for full load. This is the operating point which is of most interest for knocking and misfiring. Using it for half load is possible, since the controller will have to react on different values either way. The model shall be used for validation on MVEMs and the HiL system.

B. Fundamentals

Parameters of different operating points are provided by the LEANOX controller. With this controller, the operating values for a certain emission level are already present. Since the parameter set for the controller is determined by measurement equipment, the desired NO_x value for the operating point is present. This controller can also be used for initial implementation of an emissions model. Different influences on emissions are already implemented and taken care of by the parameter files.

For a gas mixing model, the requirements are the same. The lack of measurement data available means no suitable model can be derived from it. Because of the many gases that can be used as fuel, a large data set would be needed. Implementing the model with such precise data, would mean a very realistic implementation of the gas model, but makes the realization of such a model, with the given requirements of easily changing the parameters, obsolete. Implementing a gas model is therefore not further pursued.

C. Adopting the formulas

For the emission model, the equations from [8] and [9] are used as starting point. Variables of interest

in those equations is burnt zone temperature. One factor of the emissions is the combustion temperature which determines the amount of NO_x in the exhaust gas. The LEANOX controller calculates the desired boost pressure. In a lean burn engine these values are dependent on each other. Lean burn engines have an excess amount of air in the mixture which increases the efficiency of the engine. Increasing the power is achieved by increasing the mixture volume inside the combustion chamber through turbo charging. By increasing the mixture volume inside the combustion chamber, more air is also forced into the cylinder, which in turn reduces the temperature during the combustion process. With less combustion temperature, the amount of NO_x is decreased.

It is therefore assumed, that the behaviour of increasing the boost pressure has a similar behaviour on emissions as the burnt zone temperature. Since the connection between boost pressure and temperature is not known, measurement data are used to estimate those values. Other variables from the equations are pooled with the variables that shall be determined. The equation used for the NO_x model is given with

$$f(x) = v_1 e^{-v_2 x} \quad (6)$$

where x represents the boost pressure in the intake manifold. With this equation, the values for v_1 and v_2 can be calculated based on the measurement data.

Measurement data are from a mix of different engine setups. Since a large data pool is available, certain criteria for use able measurements are determined. Because the LEANOX controller estimates the boost pressure, measurements with varying boost pressures are important. Another aspect, that will influence the emissions is ignition timing. Resulting in measurements with varying ignition timing and boost pressures being used. Measurements are also taken from a single engine series, to rule out differences caused by other parameters. This engine type has an output power of around 1500 kW. Since the engine setup varies between those measurements, normalizing them is necessary. This also makes it easier to compare the data in future implementations.

With the normalized data, the LSQ function is used to determine parameters for the equation in 6. One example for these measurement is depicted in Figure 3 where the influence of boost pressure and ignition timing is depicted. Over the different results from the measurement data, an averaged value is taken. This will replicate the NO_x values more accurately since not one measurement will dominate.

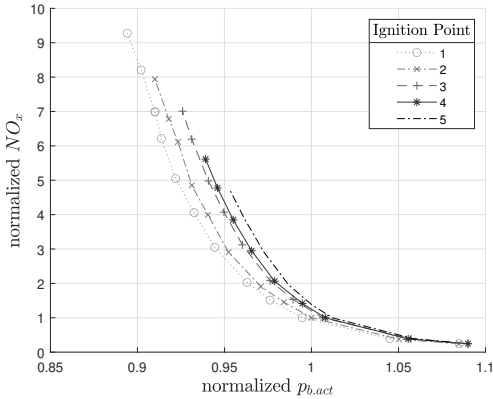


Fig. 3. Example of used measurement data for the LSQ estimation.

IV. RESULTS

Through curve fitting on the data depicted in Figure 3, the values are implemented giving

$$NOx_{Est} = \frac{NOx_{Nom}}{2.141 \cdot 10^{10} e^{-23.787 Boost_Factor}} \quad (7)$$

as final equation for the model. The values for v_1 and v_2 are already implemented. Because the data are normalized values, the return value from the equation is divided with the targeted NO_x values. Since regulations state the limits for NO_x emissions in $mg\ m^{-3}$, the values are calculated as such.

Also included is a boost factor, which is given by the relation between the reference and actual boost pressure in

$$Boost_Factor = \frac{p_{b.ref}}{p_{b.act}} \quad (8)$$

where the actual boost pressure is divided by the reference pressure to replicate the behaviour. With consideration that an increased actual boost pressure will lower the amount of NO_x in the exhaust. A decreased boost pressure will therefore increase the amount of NO_x emissions. In combination with the exponential function, the targeted value divided by the normalized result of the boost function returns the estimated NO_x values.

Initial validation of this model is done using a reduced LEANOX model. In the reduced model only the implementation of the NO_x model is tested. Quick adaptations can be implemented since little computational effort is needed. Results of the implementation can be seen in Figure 4. Where the behaviour of the NO_x model can be seen in comparison to the measurement data. The behaviour is smoother as the measurement data, but the curve shows the same behaviour. A larger error occurs, when the boost pressure is deviating far from the reference boost pressure.

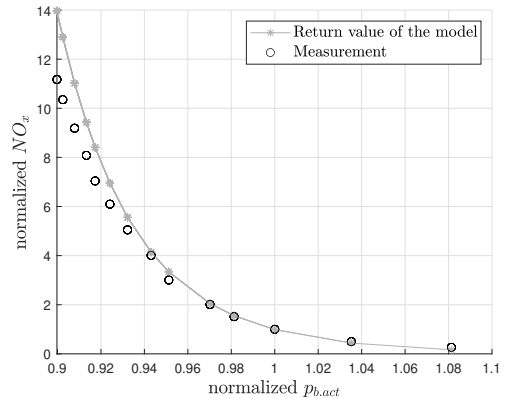


Fig. 4. Validation of the LSQ Curve fit equation in the reduced LEANOX model

Since the desired behaviour is demonstrated by the reduced engine model, the emission model is implemented in the MVEM. In the MVEM more computational power is needed since more processes

of the engine are simulated. This also means, a more realistic simulation is achieved. In the MVEM implementation actual simulated values are committed to the emission model. By manipulating these different input values, the model is again validated against measurement data, as seen in Figure 5. Where a satisfactory result can be seen because of the similar behaviour.

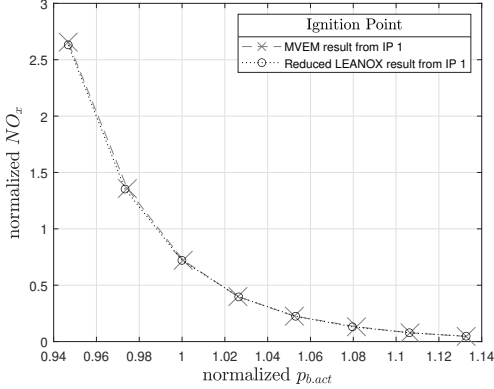


Fig. 5. Validation of the reduced LEANOX model against the MVEM.

In the final validation, the NO_x model is implemented on the HiL system. Since the HiL system replicates a real world system in a more accurate manner, some changes on the model are needed. By replicating a real engine, the connection to the PLC is realized via controller area network (CAN) bus. Over the bus link the O_2 value of the emissions and the parts per million (ppm) value of the NO_x are transmitted. Since the initial model was created with $mg\ m^{-3}$ as output, some conversion is needed.

For the conversion, the amount of O_2 is estimated with the air to fuel ratio given by

$$x_{O_2} \approx 0.21 \frac{\lambda - 1}{\lambda}. \quad (9)$$

where λ is the air to fuel ratio. The 0.21 represents the approximate amount of O_2 in the atmosphere.

Also the difference between wet and dry measurement is of important. The NO_x sensor returns values

of wet measured emissions. Translating these values is done with the equation

$$c_{dry} = c_{wet} \frac{1}{1 - H_2O} \quad (10)$$

where H_2O is the percentage of water in the measured sample. To calculate the humidity in exhaust gas following equation

$$H_2O \approx 0.1966 - 0.8953 O_2 \quad (11)$$

is used from [10], where the conversion from wet O_2 to dry O_2 is made. Since it is also not measured with the NO_x sensor, an estimation of humidity content in exhaust gas is made.

In TA Luft [11] the reference value for oxygen in exhaust gases for gas engines is given with five percent, together with the measured or calculated O_2 level the correction factor is calculated. It is given by

$$O_{2,corr} = \frac{0.21 - O_{2,ref}}{0.21 - O_{2,meas}} \quad (12)$$

for the O_2 value. Together with equation

$$c_{mg} = c_{ppm} \frac{MW}{MV} O_{2,corr} \quad (13)$$

the conversion from ppm to $mg\ m^{-3}$ is possible. Values for molecular weight (MW) and molar volume (MV) are known, with following values:

- molecular weight for $NO_2 = 46\ g\ mol^{-1}$
- molar volume = $22.4\ m^3\ mol^{-1}$

The weight of NO_x is estimated by NO_2 . In lean burn engines the amount of NO_2 is higher in exhaust gas composition [12].

The final conversion is given by

$$NO_{X,ppm} = \frac{NO_{X,mg} 22.4 (0.21 - O_{2,meas})}{46 (0.21 - O_{2,ref})} \quad (14)$$

and used for the implementation into the HiL system which returns the ppm.

The HiL system also shows a great result with the given model, which can be seen in Figure 6. In this Figure, a closed loop emissions controller reacts on a change in charge temperature by increasing boost pressure.

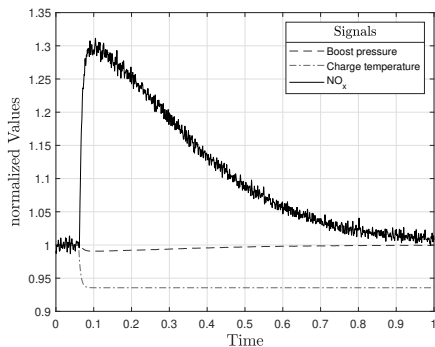


Fig. 6. Result of a real controller on the NO_x model

V. CONCLUSION

A NO_x model for controller validation is created. The behaviour of the emissions is replicated in a satisfactory manner and on different implementations of validation tools. It is also easy to adapt to different emission standards. With the correct parameter file, different engines are also implemented in a satisfactory manner.

The implementation of the gas mixing model, can not fulfill the desired requirements. It should be easy to implement and switch between different operating modes. Because of this the gas mixing is not implemented in the emissions model.

Further investigation is also needed for different engine types. A difference behaviour for different engine types can be expected and eventual different values for the model are needed. The implementation of a gas mixing model is also still outstanding.

REFERENCES

- [1] Herdin Guenther, "Standesanalyse des gasmotors im vergleich zu den zukunftsstechniken (brennstoffzellen und mikroturbine) bei der nutzung von aus biomasse gewonnenen kraftstoffen," Germany, 2002. [Online]. Available: <https://docplayer.org/8324668-Jenbacher-ag-dr-g-r-herdin.html>
- [2] R. van Basshuysen and F. Schäfer, Eds., *Handbuch Verbrennungsmotor: Grundlagen, Komponenten, Systeme, Perspektiven*, 8th ed., ser. SpringerLink Bücher. Wiesbaden: Springer Vieweg, 2017.

- [3] J. Li, C.-M. Gong, Y. Su, H.-L. Dou, and X.-J. Liu, "Effect of injection and ignition timings on performance and emissions from a spark-ignition engine fueled with methanol," *Fuel*, vol. 89, no. 12, pp. 3919–3925, 2010.
- [4] E. HU, Z. HUANG, B. LIU, J. ZHENG, and X. GU, "Experimental study on combustion characteristics of a spark-ignition engine fueled with natural gas-hydrogen blends combining with egr," *International Journal of Hydrogen Energy*, vol. 34, no. 2, pp. 1035–1044, 2009.
- [5] G. Quirchmayr, F. Gruber, and S. Knippitsch, "Device for regulating the proportion of combustion air to gaz," AT Patent EP0 259 382 (A1), 1988.
- [6] L. Guzzella, *Introduction to Modeling and Control of Internal Combustion Engine Systems*, ser. Springer eBook Collection Engineering. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010.
- [7] ZELDVICH Y. B., "The oxidation of nitrogen in combustion and explosions," *J. Acta Physicochimica*, vol. 21, p. 577, 1946.
- [8] F. Maroteaux and C. Saad, "Combined mean value engine model and crank angle resolved in-cylinder modeling with nox emissions model for real-time diesel engine simulations at high engine speed," *Energy*, vol. 88, pp. 515–527, 2015.
- [9] C. Quérel, O. Grondin, and C. Letellier, "Semi-physical mean-value nox model for diesel engine control," *Control Engineering Practice*, vol. 40, pp. 27–44, 2015.
- [10] VDMA, "Vdma 6299 - methoden zur überwachung der emissionen von verbrennungsmotoranlagen," 2019.
- [11] Bundesministerium für Umwelt, Naturschutz und Reaktorsicherheit, and www.bmu.de, "First general administrative regulation pertaining the federal immission control act," 2002.
- [12] G. P. Merker, Ed., *Combustion Engines Development: Mixture Formation, Combustion, Emissions and Simulation*, ser. SpringerLink Bücher. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012.



Cornee Sanner is student in the Department of Mechatronics at MCI Innsbruck/Austria and is a system developer at INNIO Jenbacher GmbH & Co OG.

Development of a Portable Hand Rehabilitation Device for Sensory Motor Training

Sebastian Seewald, Yeongmi Kim (supervisor)

Abstract—Millions of people suffer from stroke worldwide and the number is increasing annually. A patient's effective therapy time is severely restricted due to limited staff and robotic-assisted equipment. Especially, available robotic hand rehabilitation devices are bulky and expensive and therefore limited in access, also after clinical treatment. Therefore, the development of a lightweight, portable and cost-effective robotic hand rehabilitation device with advanced functionalities beyond simple hand movements is processed. A kinematic mechanism driven by a DC motor is developed that enables natural opening and closing of the hand in a portable way. A sensory motor training system with embedded force sensors and vibration motors are implemented in the area of fingertip pods, which transmit important data as well as stimulating feedback to the user. Measurements and tests with a prototype results in a range of motion (ROM) angle of 94° with respect to the metacarpophalangeal (MCP) joint of the hand. The maximum active support force is 59.4 N. The entire movement mechanism is compact and located within the hand by using a remote center of motion (RCM) tele-mechanism linkage. Its optimally supporting the degree of restriction and with the measured data a possible course of therapy can be medically assessed. For a single finger a force input of 0.5 to 50 N could be measured and a vibro-tactile feedback between 20 to 300 Hz could be generated. The sensory training module shows, that neurological limitations can be measured, trained and improved.

Index Terms—Hand-rehabilitation, Stroke, Haptic feedback, Sensory motor training, RCM mechanism

I. INTRODUCTION

MANY of the survivors from a stroke are affected by motoric impairments. Most stroke

S. Seewald is with the Department of Mechatronics, MCI, Innsbruck, Austria, e-mail: s.seewald@mci4me.at.

patients require rehabilitation therapy. This can help them recover from their lost physical and cognitive function. Rehabilitation treatments need to start as soon as possible after a stroke, with the maximum recovery occurring during the first three to six months. While frequent training is a key for the recovering process. Rehabilitation is usually done in specialized centers and robotic devices assisting this are extremely expensive and not available in sufficient numbers. Clinical studies showed the effectiveness of robot-assisted rehabilitation to relearn the cognitive and motor functions and highlights the need for robot-assisted rehabilitation applications in addition to the classic clinical therapy exercises [1] [2].

Especially, hand rehabilitation is necessary to perform functions and tasks in daily life scenarios. Basic activities such as eating, drinking, moving objects or writing are severely impaired depending on the course of a stroke. Many robot-assisted hand rehabilitation devices are very large, not easy to operate and usually too expensive for a possible home rehabilitation. Generally, specific limitations can only be relearned to a limited extent with these devices. Support depending on the movement impairment, motor training or a combination of motor and sensory training, for example, are only partially addressed or could be only executed with stationary hardware. Thus, the basic goal is to develop applications that can cover these functions in one small device. Furthermore, it should also be possible to use them in addition to their use in hospitals or rehabilitation centres. Properties such as small size, portability, easiness to use, cost-effective and measurement data for therapeutic or clinical interventions by specialists

are the focus.

The hand is a masterpiece of biomechanics. It plays an important role in the development of the brain's motor functions. The hand is one of the most complex structures in the human body due to its compact size, large executable degree of freedom (DOF), and range of motion (ROM). Therefore, researchers are trying to copy the hand movement as realistically as possible to be able to perform end-effector manipulations [3] with robots. On the other hand, the aim is to use the existing knowledge of human hand anatomy to reproduce the movement of the hand as naturally as possible using so-called exoskeletons [4], [5], [6] so that the motor movements of the individual limbs are not restricted in the case of corresponding support tasks. Size requirements, weight, ergonomics, rehabilitation tasks, actuators or sensors are all due to the complex anatomy and biomechanics of the hand the major challenges in developing robot-assisting rehabilitation devices.

II. SYSTEM DESIGN

A. Background

Mechanisms are often used that can reproduce the natural movement of the hand but are not compact due to their kinematics like [7], [8] and [9]. Concerning the anatomy of the hand the finger consists of three phalanges and joints. The most proximal is the metacarpophalangeal (MCP) joint, which has an flexion and extension ROM of approx. 90° and 40° from the natural rest. The next joint is named the proximal interphalangeal (PIP) joint and possesses a flexion movement of around 130° . The hardware must be designed in a way that both joints can be flexed simultaneously up to a maximum of their respective ROM. In summary, the device is required to possess a mechanism that allows manipulation of the MCP's and PIP's angular position respectively to induce therapeutical effects to the specific muscles. But also the movement of the wrist should not be limited in its range of motion.

B. Mechanical Design of the Hand Rehab Device

In order to allow actuation of two joints simultaneously a specific configuration of links is chosen often referred to a remote centre of motion (RCM) mechanism. This implies that additional to an actively actuated joint another joint's rotation at a different location from the centre of rotation is carried out passively.

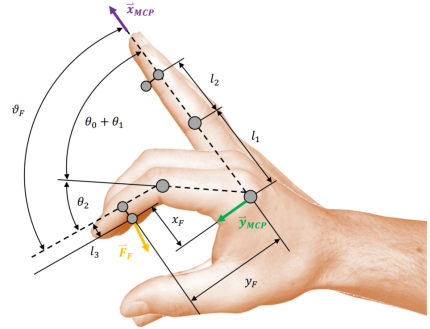


Fig. 1. Schematic representation of the finger kinematics, where $\theta_0 = 0$ when all finger links are in a straight line and equal to vector \vec{x}_{MCP} .

To calculate the finger position coordinates (x_F , y_F and θ_F) the following equations (1)-(3) are used.

$$x_F = l_1 \cos(\theta_0 + \theta_1) + l_2 \cos(\theta_1 + \theta_2) - l_3 \sin(\theta_1 + \theta_2) \quad (1)$$

$$y_F = l_1 \sin(\theta_0 + \theta_1) + l_2 \sin(\theta_1 + \theta_2) - l_3 \cos(\theta_1 + \theta_2) \quad (2)$$

$$\Theta_F = \sum_{i=0}^2 \theta_i \quad (3)$$

As can be seen in Fig. 2 some preliminary research works showed that there are angular relationships between MCP and PIP joints during hand movements. For the development of the portable hand rehabilitation device (PHRD) it can be assumed that

there is a constant coupling (marked as continuous line in the graph) between MCP and PIP joint angle. Same like in [10] the value of this constant coupling is chosen in the range of $n_{12} \in [1.25 \ 1.75]$.

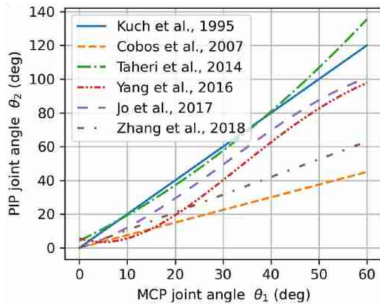


Fig. 2. Angular relationship between PIP and MCP joints [10].

Therefore, the initial two DOF kinematic model can be reduced to a one DOF model as can be seen in the next equation (4).

$$\theta_2 = n_{12} \cdot \theta_1 \quad (4)$$

When opening and closing the hand, a translational displacements also occur with respect to the MCP joint. In order to compensate this, a tele-mechanism is being developed, which allows a translational displacement of up to 50 mm. The result is expected to be a mechanism which shows great similarity to the motion pattern of a real human finger excluding the movement and rotation of the DIP joint. With Autodesk Inventor® a RCM tele-mechanism is implemented and simulated accordingly to fulfill the requirements for a natural opening and closing motion of the hand. With the dimensions of a reference hand, the raw basic mechanism is implemented accordingly in a functional and production-oriented manner. The final prototype, which consists of several sub components is shown in Fig. 3. It could be divided in three main parts:

- **Upper inner housing** contains the stable connection to the RCM tele-mechanism, whereby a single-stage gear ratio ($i = 5$) and the modified DC motor (HV7346MG, AM Racing) with a torque of 4.6 Nm at the motor shaft are fixed and mounted. Furthermore, a high-precision position sensor (3382H-1-103, Bourns) is installed axially on the final stage of the transmission in order to be able to control the angular position of the RCM tele-mechanism. This section is the basic structure of the unit and serves as the connection point for all other components.

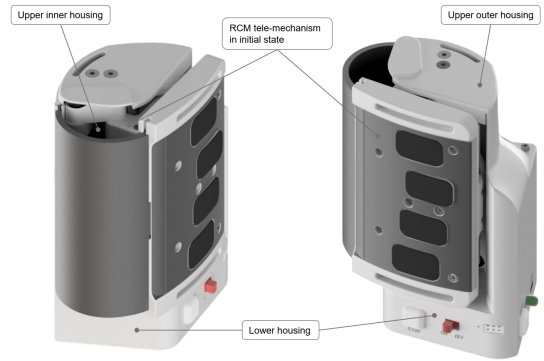


Fig. 3. Final prototype of the PHRD with a sensory training module in initial position of the RCM tele-mechanism.

- The **upper outer housing** serves mainly as an optical cover and as an ergonomic support surface for the thumb and the rest of the hand. Also this part represents as an interface for the fixation of the hand to the device. A LiPo battery (7.4 V and 650 mAh, Tattu) is installed, rotated 90° to the RCM mechanism, which provides the necessary power supply for the printed circuit board (PCB), actuators and sensors.
- The **lower housing** is fixed to the upper inner housing with two screws. Essentially, it contains all the electronic components such as PCB with microcontroller (Arduino Nano 33 BLE), inertial measurement unit (IMU) (LSM9DS1), volt-

age regulator (TLV1117, Texas Instruments), motor driver (TLE6209R, Infineon), multiplexer switch IC (TMUX1208-Q1, Texas Instruments), bluetooth module, charging port, data port, buttons and switches. Furthermore, this module serves as the central interface to the external components such as sensors and actuators.

C. Sensory Training System

A sensory training system (STS) with embedded force sensitive resistor (FSR) sensors (FSR04CE, Ohmite) and vibration motors (316040001, Seeed) are implemented in the area of the fingertips, which transmit data as well as haptic feedback to the user. The design of this advanced functionality is illustrated in Fig. 4.

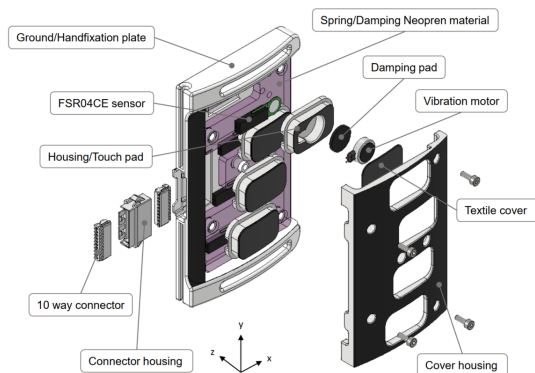


Fig. 4. Explosion view of the STS module.

The keypad contains an embedded vibration motor that provides direct contact with the finger in order to perceive the vibro-tactile stimuli. Furthermore, the keypad element is guided by two guidance domes and the damping neoprene material generates an offset force of 0.5 N to prevent involuntary actuation during operation.

Haptic feedback can be employed to measure reaction times after a stimulation with a single vibration motor or use it in a playful game content. 3D

printed stereolithography (SLA) components are used to achieve sufficient accuracy and fit to the individual components. Furthermore, it should be ensured that the force focus domes always interact exactly with the active surface of the FSR sensors.

D. Electronics

The FSR sensor uses the principle that resistance can be changed by pressure at both sides of a contact surface. A circular FSR sensor with an active area with a diameter of 5.6 mm is used and can detect force inputs from 0.2 to 50 N. In order to obtain reference values for the later software implementation, a reference curve is determined. A realistic test setup is implemented, where different norm weights are applied to the sensor. The supplied voltage values are recorded with a Arduino test program and a Excel data logger. The measurement data and the resulting characteristic curve is shown in Fig. 5.

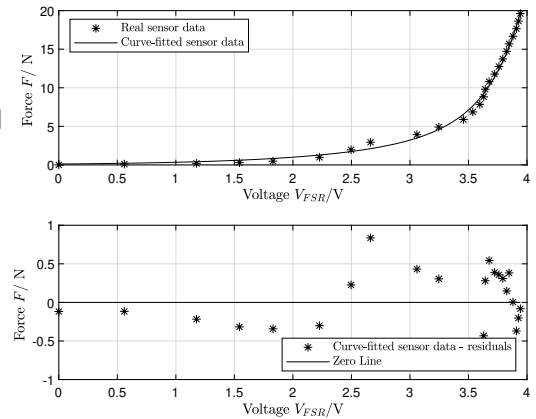


Fig. 5. Reference curve of the used FSR04CE sensor between force impact to voltage output. The sensor data are fitted accordingly using a polynomial function.

The RCM tele-mechanism is driven by a modified DC motor and gear stage. In order to obtain a more precise and reliable position determination of the moving part, the servo motor is converted to a

DC motor and the position value is realised via an externally installed 12 mm rotary position sensor.

The Arduino Nano 33 BLE is compact (45x18)mm and provides a bluetooth module for communication. Also a embedded 9 axis IMU to collecting measurement data and informations about orientation, movement and vibrations of the device is implemented. By using a current sensor module (INA219, Texas Instruments) and a built-in shunt resistor, the current consumption is measured during the movement of the motor and RCM tele-mechanism and serves as an additional control stage or as a safety component when a maximum torque consumption is reached, whereby the system is stopped. A schematic illustration of the entired electronics is shown in Fig. 6.

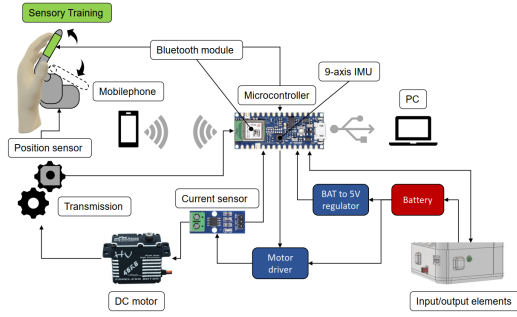


Fig. 6. Schematic circuit of the entired implemented electronics in the PHRD.

The whole electronics is assembled together on a PCB which is mounted in the lower housing.

E. Software

Subsequently, movement limitations can be determined by means of current measurement, as

$$T_m = k_m \cdot i_m. \quad (5)$$

is valid and this provides direct information about the existing torque that is exerted on the hand and thus enables safety-relevant switch-off limits during

the operation. Furthermore, with the same information the maximum ROM of an user can be calculated with the help of the known position when a threshold is exceeded. The maximum possible opening and closing angle is limited by software using the mentioned position sensor. For safety reasons, all angles outside this range are mechanically limited by the housing.

For the basic sensory training function, a sequence is implemented, which is shown in Fig. 7. Two important indicators are identified: Controllable vibrotactile feedback is used to determine the reaction time until the sensor-motor training system is actuated in the area of the fingertips. If the threshold is exceeded, the feedback is interrupted. Furthermore, the temporal course of the force impact as well as the maximum applied force are determined and stored accordingly.

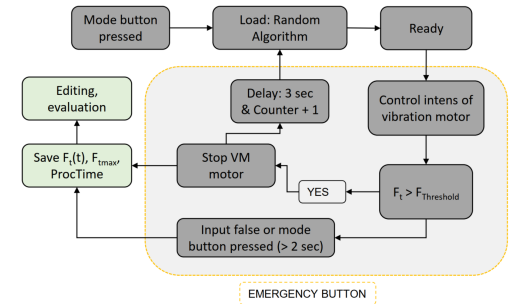


Fig. 7. Procedure of the first stage of a sensory motor training.

III. RESULTS

A. Study with FH Gesundheit

In a cooperation with the FH Gesundheit, the first PHRD prototype is tested and analysed by 7 trainee occupational therapists and corresponding evaluation scales are developed. In a period from 06.05. to 13.06.2022 and in a total of three face-to-face dates in the MCI laboratory, the students were introduced with the prototype and its functions and tasks were

explained. The device was then be tested and evaluated by the students themselves. In a final report [11] the findings and results are presented. A system usability scale (SUS) as well as a usefulness, satisfaction and ease of use (USE) questionnaire are filled in and elaborated. Aspects from a therapist's point of view, such as effectiveness in use, are assessed by means of a score with 7 digits from *strongly disagree* (1) to *strongly agree* (7). The results are shown in Table I.

TABLE I
RESULTS OF THE USE QUESTIONNAIRE [11]

Description	Value
Usefulness	6
Ease of Use	7
Ease of Learning	7
Satisfaction	5

B. DC Motor Current Measurement

A current sensor circuit is integrated which makes it possible to measure the current consumption of the DC motor.

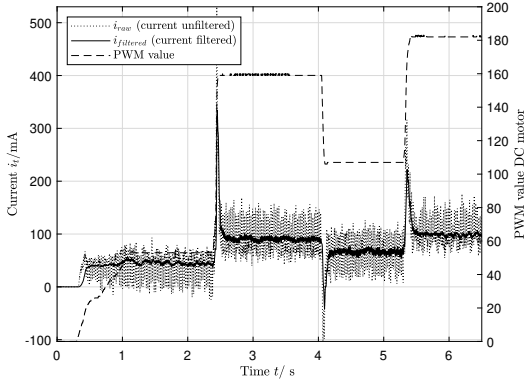


Fig. 8. Measurement data of the current sensor INA219 during motor operation with different PWM values to evaluate the quality of $i_{filtered}$ by using a Kalman filter function.

Since the linear correlation to the existing torque applies to a DC motor, the measured current can also

be used to identify the actual torque effecting on the hand through the mechanism. In order to obtain accurate measured values, the necessary application is tested and optimised. Fig. 8 shows the course of the motor current over time. The current consumption with different pulse width modulation (PWM) value of the motor is shown. By means of a Kalman filter function, the system noise is suppressed and the exact current value is determined. The accuracy is 2 mA compared to the reference measurement using a amperemeter.

C. Sensory Training System

To show the results of the STS, a procedure is implemented using the Arduino integrated development environment (IDE). It measures the time between stimuli input via the vibration motor on one of the four pressure pads and reaching the threshold, for example, selected with 7 N.

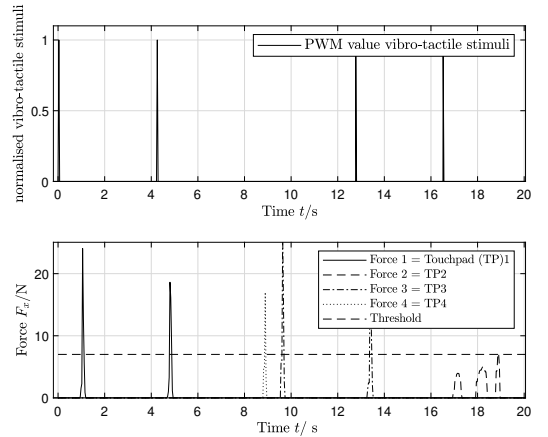


Fig. 9. Measurement data of the sensory training with four vibro-tactile stimuli to measure the reaction time t_r and force input F_t .

Afterwards, the maximum force impact when pressing the touch pad will be stored. After a successful stimuli-reaction procedure a new random stimuli

is set to one of the four pressure pads. As can be seen in Fig. 9 the normalised force stimulation value (logical value = 1) and the corresponding force impact through the fingers are shown over time.

Also with several measurements with different test persons in different age and partially proven pre-existing diseases (Parkinson, former hand injury, etc.), significant and clearly assignable measurement results can be achieved, which can be shown, for example, in the ability to react and the impact of force. The vibro-tactile stimulation could be applied to the four fingertips with a frequency up to 300 Hz. The haptic feedback is precisely introduced by the decoupled pressure pads and allows frequencies down to 20 Hz to be perceived.

D. Final Prototype

The final prototype (Fig. 10) consists of the portable handheld rehabilitation device and the sensory training system.

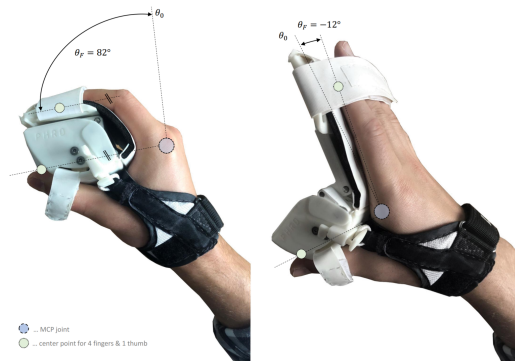


Fig. 10. Real final prototype of the PHRD with the additional finger and hand fixation in the maximum closed and open state.

The sensory system can be easily exchanged in a modular way and allows the prototype to be used in a variety of ways. A flexible fabric cover between the moveable STS housing and stationary housing is clipped in for safety reasons so that the hand is not

hurt between some moveable rigid components or the gear stage. It is also used as a guidance of the wires from the sensors and actuators from the module to the electronic unit in the lower housing.

The dimensions of the final prototype in close-state position without finger, thumb and hand fixation is (HxBxD) (126.3x86.9x63.7)mm. The weight of the basic version, which is shown in Fig. 10 is 455 g. The developed extended sensory training module weights 55 g. From the resting position of the hand marked with θ_0 , an MCP angle of 82° is possible when the fingers are closed. When the mechanism is fully opened, an angle of $\theta_F = -12^\circ$ is possible with response to the resting position. This allows a total 94° of movement in the xy-plane.

IV. CONCLUSION

The aim of this paper is to design a hand rehabilitation device enables rehabilitation of both - motor and sensory function training of the hand. The design supports high forces, a natural ROM and is a cost-effective (262,61€ for basic version) solution compared to similar devices on the market or in the research field.

The results of the force calculation shows that maximum opening and closing forces up to 59.4 N are possible for the hand. According to [7] the maximum grasping forces for activities of daily living range from 20 N to 40 N. Therefore, the device shows the potential to provide sufficient forces applied to the hand to train activities of daily living.

The ROM analysis of the developed system yields average bending angles of 94° at the MCP joints and 40° at the PIP joints, which is a higher ROM than [9] (49.9° bending angle at the MCP joint). However, with the achieved ROM, the user is able to train a natural opening and closing of the hand and furthermore with a impedance control from [12] virtual objects can be grasped and moved with in a virtual reality (VR) environment and necessary

haptic feedback can be executed to the user.

The weight of the PHRD in the basic version is 455 g. The sensory training system itself weights 55 g. That is about the same, compared to devices developed in [10] and [5], but this comes from the fact, that all electronic elements, are mounted directly onto the device and the whole setup is fully portable. A optimization of the final prototype, for example, by using lighter materials or adjust some structural parts the weight can still be reduced by a few percent.

ACKNOWLEDGMENT

I would like to thank FH-Prof. Yeongmi Kim PhD. for the great support during the entire work. A big thank you also goes to Thomas Gadner and Stephan Jäkel, who helped me with the implementation of electronic subject areas. Further thanks go to Simon Deiser for the excellent cooperation of the overall interdisciplinary project work.

REFERENCES

- [1] N. G. Kutner, R. Zhang, A. J. Butler, S. L. Wolf, and J. L. Alberts, "Quality-of-life change associated with robotic-assisted therapy to improve hand motor function in patients with subacute stroke: a randomized clinical trial," *Physical therapy*, vol. 90, no. 4, pp. 493–504, 2010.
- [2] C. D. Takahashi, L. Der-Yeghian, V. Le, R. R. Motiwala, and S. C. Cramer, "Robot-based hand motor therapy after stroke," *Brain : a journal of neurology*, vol. 131, no. Pt 2, pp. 425–437, 2008.
- [3] Shadow Robot Company, "Dexterous hand series: The world's most dexterous humanoid robot hands," 2022. [Online]. Available: <https://www.shadowrobot.com/dexterous-hand-series/>
- [4] T. Bützer, O. Lamercy, J. Arata, and R. Gassert, "Fully wearable actuated soft exoskeleton for grasping assistance in everyday activities," *Soft robotics*, vol. 8, no. 2, pp. 128–143, 2021.
- [5] A. Mandeljc, A. Rajhard, M. Muni, and R. Kamnik, "Robotic device for out-of-clinic post-stroke hand rehabilitation," *Applied Sciences*, vol. 12, no. 3, p. 1092, 2022.
- [6] I. Jo and J. Bae, "Design and control of a wearable and force-controllable hand exoskeleton system," *Mechatronics*, vol. 41, pp. 90–101, 2017.
- [7] M. Decker and Y. Kim, "A hand exoskeleton device for robot assisted sensory-motor training after stroke," in *IEEE World Haptics Conference (WHC 2017)*. Piscataway, NJ: IEEE, 2017, pp. 436–441.
- [8] D. Wang, Q. Meng, Q. Meng, X. Li, and H. Yu, "Design and development of a portable exoskeleton for hand rehabilitation," *IEEE transactions on neural systems and rehabilitation engineering : a publication of the IEEE Engineering in Medicine and Biology Society*, vol. 26, no. 12, pp. 2376–2386, 2018.
- [9] O. Sandoval-Gonzalez, J. Jacinto-Villegas, I. Herrera-Aguilar, O. Portillo-Rodriguez, P. Tripicchio, M. Hernandez-Ramos, A. Flores-Cuautle, and C. Avizzano, "Design and development of a hand exoskeleton robot for active and passive rehabilitation," *International Journal of Advanced Robotic Systems*, vol. 13, no. 2, p. 66, 2016.
- [10] R. Rätz, F. Conti, R. M. Müri, and L. Marchal-Crespo, "A novel clinical-driven design for robotic hand rehabilitation: Combining sensory training, effortless setup, and large range of motion in a palmar device," *Frontiers in Neurobotics*, vol. 15, p. 748196, 2021.
- [11] A. Brunner, J. Christandl, S. Hohenwarter, B. Mayr, Lena; Rohregger, A. Stockinger, and J. Zauchner, "Prototype Testing for NeuroRehab-Devices(Usability)," FH-Bachelor-Studiengang Ergotherapie, 2022, FH Gesundheit.
- [12] S. Deiser, "Admittance control and adaptive assistance rehabilitation for a portable hand rehabilitation device," Master Thesis, 2022, Management Center Innsbruck.



Sebastian Seewald studies Mechatronics and Smart Technologies at the MCI Innsbruck/Austria with focus on mechanical engineering. He also works as a development engineer at Hirschmann Automotive GmbH where he is responsible for development of special connectors and wire harnesses in the automotive environment.

Deep Learning-based Condition Monitoring Framework for Rolling-Element Bearings

Hansi Strobel, Franz-Josef Falkner (supervisor)

Abstract—In industrial settings, almost half of rotating machinery failures occur due to rolling-element bearing degradation. As a result, accurately estimating the internal condition of bearings is crucial to ensure reliable and safe operation. In light of the recently proven potential of the transformer architecture in natural language processing (NLP) and digital image processing (DIP) tasks, this thesis proposes a novel, transformer-based predictive maintenance framework to discern the internal condition of rolling-element bearings. The proposed deep learning architecture utilizes short-time Fourier transform (STFT) to construct spectrogram images and correlates their intrinsic time-frequency characteristics with the accumulated operating hours. As a result, the architecture is able to model the temporal evolution of the vibration patterns. The performance of the framework was tested on two popular tasks: degradation trend modeling and bearing fault classification. Degradation modeling experiments delivered good results on small datasets, with improved performance as datasets grew larger. The inclusion of the operating time token significantly increased the effectiveness of the architecture in capturing vibration-temporal patterns. During fault classification, the architecture displayed robust predictions, even in noisy environments.

Index Terms—Deep Learning, Rolling-Element Bearing, Predictive Maintenance, Short-Time Fourier Transform

I. INTRODUCTION

ROLLING-ELEMENT bearings see a wide range of applications in industrial rotating machinery [1]. To ensure safe and interrupt-free operation, operators traditionally rely on scheduling periodic maintenance [2], [3]. Though this so-called preventative maintenance scheme has proven effective, its disadvantages are excessive machine downtime and increased cost due to unnecessary maintenance at

fixed intervals [3]. Predictive maintenance schemes try to address these issues by analyzing the internal machine condition and scheduling maintenance as necessary [1].

Recent advances in deep learning sparked a new branch of research focused on developing heavily data-driven, deep learning-based predictive maintenance frameworks for bearings. These frameworks are used for both degradation estimation and failure mode classification while typically following a two-stage process: 1) feature extraction and 2) modeling of degradation using the extracted vibration information [4]. Feature extraction is often done by selecting relevant time-, frequency- or time-frequency-domain features. For instance, Nie et al. [3] used monotonicity and trendability criteria to select the most optimal features and applied principle component analysis (PCA) to reduce dimensionality of the feature space and derive an health indicator (HI). More recent work in the field utilized self-supervised auto encoder (AE) models to derive an HI. To derive a system-independent, optimal HI, Wang et al. [5] trained a convolutional AE to create a compressed representation of the sensor data. They subsequently fed this representation into a self-organising map (SOM) clustering algorithm to derive an HI. Both of these approaches employ an intermediate step that involves creating an HI before using it as the input to their prediction model. However, researchers also tried skipping this intermediate step altogether and directly used previously extracted features as the input to their models. Xu et al. [1] and Zu et al. [4] both utilized STFT to derive time-frequency-information and processed it in a modified SqueezeNet and long

short-term memory (LSTM)-transformer architecture, respectively.

In light of the recently proven potential of the transformer architecture in NLP and DIP tasks [6], [7], this paper proposes a novel, transformer-based predictive maintenance framework. Utilizing spectrogram images along with the accumulated operating time as the input, temporal spectrum transformer (TST) aims to model the temporal evolution of vibration patterns. Unlike most state-of-the-art frameworks, TST uses a non-recurrent architecture to model the internal condition of a bearing and, as a result, does not require continuous monitoring data during inference.

The remainder of this thesis is organized as follows: Section II briefly introduces the relevant theoretical concepts. Section III explains the methods used to investigate the effectiveness of the proposed framework. Subsequently, Section IV presents the results and Section V concludes the thesis.

All of the TST source code and the preformatted datasets are available on GitHub (www.github.com/TaiboloV2/TST).

II. THEORETICAL BACKGROUND

A. Supervised Learning

In general, most problems solved by machine learning algorithms can be formulated as a mapping function $\vec{f} : X \rightarrow Y$, that maps the features of the input space X to the desired output space Y . In traditional programming, the goal is to find this mapping function by manually specifying the mathematical relationship of input space features X , which result in the desired output Y . However, with increasing complexity of the problems at hand, manually deriving a working algorithm becomes more and more challenging. [8]

Supervised learning assumes that it is easy to obtain sufficient input data along with their solutions $(\vec{x}, \vec{y}) \in X \times Y$, where the solution \vec{y} is commonly referred to as the label [9]. Using this data, supervised learning aims to find the function $\vec{f} \in \mathcal{F}$ (model) that best describes the input-output relationship across the

training dataset $\{(\vec{x}_1, \vec{y}_1), \dots, (\vec{x}_n, \vec{y}_n)\} \in X \times Y$. The type and complexity of functions in the search space \mathcal{F} depends on the network architecture and can range from a simple linear, to a highly non-linear relationship. The terms *learning* or *training* directly refer to the aforementioned searching procedure.

A scalar-valued loss function $L(\hat{\vec{y}}, \vec{y})$ is employed to measure the disagreement of the predicted labels $\hat{\vec{y}}_i = \vec{f}(\vec{x}_i)$ and the true labels \vec{y}_i , thereby providing a means to measure the quality of a model. The objective of training therefore becomes finding a function that minimizes the loss function $L(\hat{\vec{y}}, \vec{y})$ across the entire input space X and output space Y .

B. Optimization

The previous section described the concept of learning as searching a function $f \in \mathcal{F}$ that minimizes the loss function L . Now, the search space \mathcal{F} is limited to a certain class of functions, dictated by the architecture of the neural network [8]. Taking a very simple neural network architecture, univariate linear regression Equation 1, it is obvious that the function f may be changed to any other function in \mathcal{F} by altering its learnable parameters $\vec{\theta}$.

$$f(x, \vec{\theta}) = \theta_1 x + \theta_2 = \hat{y} \quad (1)$$

Recall that we use a scalar-valued loss function $L(\hat{y}, y)$ to measure the quality of the mapping function $f : X \rightarrow Y$. Substituting Equation 1 into L yields: $L(f(x, \vec{\theta}), y)$, revealing the underlying connection to linear regression. By restricting the network architecture to only differentiable functions, it is possible to compute the gradient of the loss function $\nabla_{\vec{\theta}} L$ with respect to the learnable parameters $\vec{\theta}$. Utilizing the gradient $\nabla_{\vec{\theta}} L$ the problem is reduced to solving an optimization problem that aims to find an extremum of the loss function. In the context of most common machine learning applications, the problem is posed in a way where the goal is to minimize the loss, seeking a minimum of the loss function. Most commonly, the gradient of the loss function

is computed utilizing the backpropagation algorithm [8].

C. Neural Networks

Feedforward neural networks, also called multilayer perceptron (MLP) or fully-connected networks, are the fundamental deep learning models. The term *feedforward* alludes to the fact that information within these networks flows from the input \vec{x} via the intermediate connections to the output \hat{y} . Information only flows forward through the network and there are no feedback connections allowing for the output to be fed back into the network itself [9].

Feedforward neural networks are networks because they compose several functions $f^{(i)}$ into a single function f . Take three functions $f^{(1)}$, $f^{(2)}$ and $f^{(3)}$. Chaining them together results to $f(\vec{x}) = f^{(3)}(f^{(2)}(f^{(1)}(\vec{x})))$, a structure commonly seen in neural networks [9]. In this example, $f^{(1)}$ is the first layer of the network, $f^{(2)}$ to second and $f^{(3)}$ the third and final layer. The final layer is referred to as the *output layer*. The input \vec{x} is coined the *input layer* and layers between input and output are referred to as *hidden layers*. Figure 1 illustrates an exemplary network structure for the feedforward neural network $f(\vec{x})$ described above.

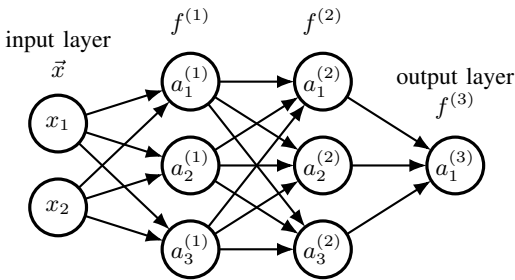


Figure 1: Diagram of MLP with two hidden layers.

III. METHODS

A. Overview

This thesis proposes a novel end-to-end prognostics framework capable of modeling rolling-element bearing degradation. Two different use-cases are proposed: degradation trend modeling and fault classification. The first of which utilizes raw vibration signals and temporal information, while the second only uses vibration signals. Figure 2 shows an overview of the framework and the flow of information within.

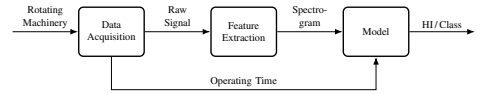


Figure 2: Prognostics framework overview.

B. Data Acquisition

In the first step, data acquisition, the raw vibration signal is measured and recorded using a vibration transducer and data acquisition (DAQ) system. Recording may either happen periodically or in a continuous fashion.

Either way, a sliding window is used to separate the raw signal into segments of $L \in \mathbb{N}$ samples. In addition to the vibration signal sample vector \vec{x}_i , the current operating time t_i is also recorded. The recordings \vec{x}_i are separated by the stride $s \in \mathbb{N}$. The stride may either be chosen in a way where all recorded sample vectors are unique ($s \geq L$) or have overlapping samples ($s < L$). This results in a total number of $N = \lfloor \frac{T f_s s}{L} \rfloor$ sample vectors, where T is the total bearing lifetime and f_s is the sampling frequency.

C. Feature Extraction

In the feature extraction step, the sample vectors \vec{x}_i of the prior step (data acquisition) are transformed into a time-frequency representation using STFT. This thesis utilizes a Hanning window function in the STFT procedure. The STFT output matrix $\mathbf{X} \in \mathbb{R}^{64 \times 64}$

shows 64 frequency bands computed at 64 different time instances. The magnitude of Fourier coefficients form the so-called *spectrogram*, which also acts as the one-channel input image to the TST model.

D. Model

The objective is to design an architecture capable of combining vibration and temporal information of bearing operation to determine the internal degradation. Introducing TST, a non-recurrent, transformer-based architecture utilising spectrogram images in conjunction with a special time token to model the vibration-temporal characteristics of rolling-element bearings (Figure 3).

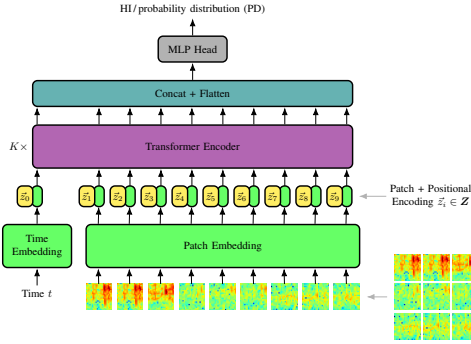


Figure 3: TST model overview.

Inspired by the vision transformer (ViT) architecture, TST reshapes a spectrogram $\mathbf{X} \in \mathbb{R}^{H \times W}$ into a sequence of flattened 2D patches $\mathbf{X}_p \in \mathbb{R}^{N \times P^2}$, where (H, W) is the resolution of the spectrogram and (P, P) is the resolution of the individual image patches. The resulting number of image patches is $N = \frac{HW}{P^2}$. Following the initial patching step, the flattened image patches are passed through the patch embedding layer where they are projected into the encoder dimension d_{model} . Analogous to the image patches, the accumulated operating time t is passed through its own time embedding layer, leading to a total $N + 1$ input tokens being passed into the transformer encoder layer(s). It must be noted that

including the time token is optional and depends on the application. Next, the column vectors \bar{z}_i of the positional encoding matrix \mathbf{Z} are added. All patches are then forwarded through K encoder layers and subsequently concatenated and flattened to a vector of size $(N + 1)d_{model}$. Feedforward layers in the MLP head either produce a scalar-valued HI or a PD across classes.

The implementation of the proposed architecture was done using Python 3.9.13 and PyTorch 2.0.0.

E. Degradation Modeling

The first use-case is degradation modeling, which was treated as a supervised regression task. Data labels were generated by employing linear degradation labeling. Linear labeling assumes that bearing degradation is a process which occurs at a constant rate over the entire lifetime of the rolling-element bearing. With the health indicator label $HI_i \in [0, 1]$ ranging from 1 (healthy) to 0 (end of life), the labeling process boils down to a line connecting the start and end of bearing operation.

The validation of the degradation modeling ability of TST is done using the FEMTO [10] condition 1 (Table I) and XJTU-SY [11] condition 1 and 2 (Table II) accelerated run-to-failure bearing datasets.

Table I: FEMTO bearings and operating conditions.

Operating Condition	Dataset	Mode of Failure	
Condition 1: 1800 rpm 4 kN	Training	Bearing1_1	-
		Bearing1_2	-
	Testing	Bearing1_3	-
		Bearing1_4	-
		Bearing1_5	-
		Bearing1_6	-
		Bearing1_7	-

F. Fault Classification

To assess TST fault classification performance using vibration and/or acoustic measurements, a new bearing test stand was constructed. The stand measures vibration and acoustic emissions of bearings with artificially induced damage (Figure 4).

Table II: XJTU-SY bearings and operating conditions.

Operating Condition		Dataset	Mode of Failure
Condition 1: 2100 rpm 12 kN	Training	Bearing1_1	Outer race
		Bearing1_2	Outer race
		Bearing1_3	Outer race
	Testing	Bearing1_4	Cage
		Bearing1_5	Inner and outer race
Condition 2: 2250 rpm 11 kN	Training	Bearing2_1	Inner race
		Bearing2_2	Outer race
		Bearing2_3	Cage
	Testing	Bearing2_4	Outer race
		Bearing2_5	Outer race

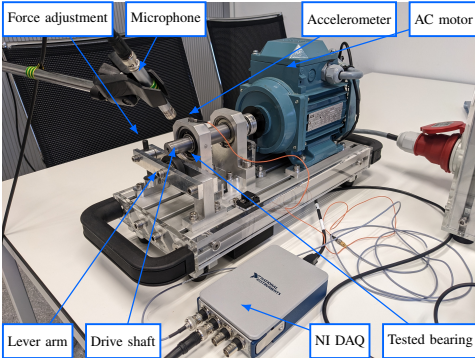


Figure 4: MCI bearing condition test bench.

While the radial load was kept constant, both the shaft speeds (300-2100 rpm) and testing environments were varied. The tested environments were:

- Laboratory with minor disturbance (LB)
- Management Center Innsbruck (MCI) workshop with a lathe (LT)
- MCI workshop with a lathe and mill (LM)

IV. RESULTS

A. Degradation Modeling

Validation of the degradation modeling capabilities of TST is conducted using the FEMTO and XJTU-SY datasets described earlier. Three variants of TST are proposed (Table III), varying in model dimension d_{model} , hidden layer dimension d_{ff} , number of encoder layers and number of attention heads. to investigate the impact of model complexity and parameter count on the resulting prediction.

Table III: TST model variants.

Model	d_{model}	d_{ff}	Layers	Heads	Parameters
TST-S	8	16	6	4	0.34M
TST-M	64	256	8	8	2.65M
TST-L	128	512	16	16	7.71M

The encoder output feeds into an MLP head (Table IV) that maps it to an HI.

Table IV: MLP head of the regression model.

Layer	Activation Function	Output Dimension
FF1	ReLU	512
FF2	ReLU	128
FF3	-	1

All model variants were trained using the adaptive moment estimation (Adam) optimization algorithm with momentum factors $\beta_1 = 0.9$ and $\beta_2 = 0.999$, as well as a weight decay of $\lambda = 0.01$. The training set was split into mini-batches of size 32, and an exponentially decaying learning rate ($\gamma = 0.9$) initialized at 1×10^{-4} was employed.

The experiments utilized the mean squared error (MSE) as the training metric. Models were trained until the test dataset MSE stopped improving for three subsequent epochs. After training has stopped, the model with the lowest total MSE across all test bearings was employed to evaluate model performance.

All model variants were trained using Bearing*_1 and Bearing*_2 of the respective datasets. Results of the initial round of experiments are shown in Table V. The highlighted result relates to the model variant that delivered the lowest prediction MSE for the respective bearing.

As the size of the training dataset is rather limited (2-5/2-3 train-test-split), generalization is key to performing well across all test bearings. Because simpler models are less prone to overfitting and tend to generalize better, it makes sense that the two less complex model variants generally outperformed the larger TST-L model.

To try and explain the results, let us look at the predicted degradation trends of the FEMTO bearings. As Figure 5 shows, the model was able to learn and

Table V: Linear degradation prediction MSE using two bearings as training data $\lambda = 0.01$.

Dataset	Bearing	TST-S	TST-M	TST-L
FEMTO	Bearing1_3	0.00621	0.00721	0.01497
FEMTO	Bearing1_4	0.01258	0.01279	0.01534
FEMTO	Bearing1_5	0.01902	0.0082	0.00837
FEMTO	Bearing1_6	0.01986	0.00665	0.00981
FEMTO	Bearing1_7	0.01633	0.01532	0.01456
Mean		0.0148	0.01003	0.01261
XJTU-SY	Bearing1_3	0.04948	0.04451	0.06485
XJTU-SY	Bearing1_4	0.21039	0.1325	0.14646
XJTU-SY	Bearing1_5	0.06261	0.04057	0.069
Mean		0.10749	0.07253	0.05918
XJTU-SY	Bearing2_3	0.00835	0.01209	0.01115
XJTU-SY	Bearing2_4	0.03487	0.03014	0.03109
XJTU-SY	Bearing2_5	0.01135	0.01389	0.03825
Mean		0.01819	0.01871	0.02683

reproduce the degradation trends of both training bearings. Furthermore, it was able to deliver good results for test bearings 1_3, 1_5 and 1_6, performing slightly worse on bearings 1_4 and 1_7. In contrast to the FEMTO dataset, the architecture performed poorly across all XJTU-SY condition 1 bearings. This is especially apparent because the model was unable to reconstruct the degradation trends of either training bearing, which is likely caused by early stopping during the training process.

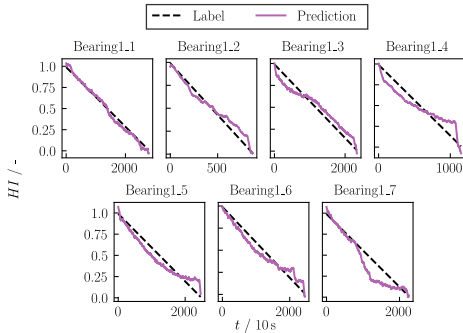


Figure 5: TST-M degradation trend of FEMTO condition 1.

As TST is a similarity matching scheme the reason for the high variance in model performance might be the quality and size of the training set. As seen in

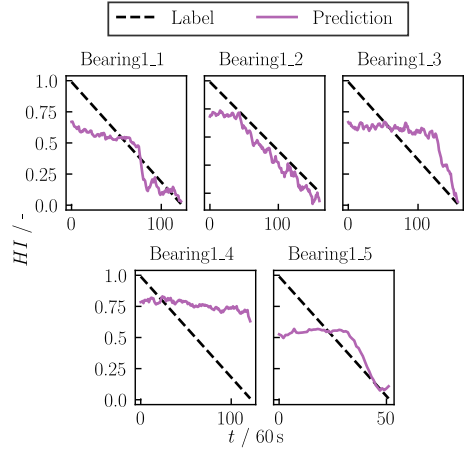


Figure 6: TST-M degradation trend of XJTU-SY condition 1.

Figure 7, the root mean square (RMS) of XJTU-SY condition 1 training bearings 1_1 and 1_2 increases about half way through their lifetime where it then plateaus for an extended amount of time. Then, right before failure, the RMS increases sharply. In contrast, the test bearings 1_3 - 1_5 experience different trends to where the RMS rises gradually and does not plateau. This difference in RMS trend might be a reason for the poor prediction accuracy.

Moreover, looking at the RMS trend it would be reasonable to believe that there exists some correlation between the RMS trend (Figure 7) and the type of physical degradation (Table II) which has caused the bearing to fail. However, no clear connection between RMS trend and bearing failure could be identified.

The earlier observations suggest that covering a wider range of degradation trends results in better model performance. For that reason, further experiments investigating the impact of larger training datasets on model performance were carried out.

The results of the experiments (Table VI) show that increasing the size of the training set leads to overall improved performance of all model variants.

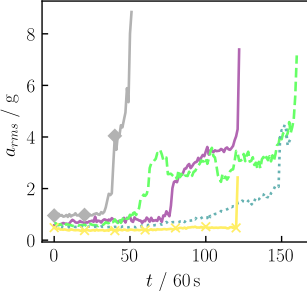
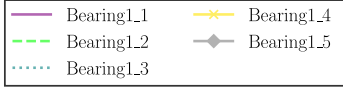


Figure 7: Acceleration signal RMS of XJTU-SY condition 1 bearing dataset.

Especially the MSE of the largest variant (TST-L) decreased significantly in most scenarios. This coincides well with the general notion that larger models tend to scale well with larger datasets.

Table VI: Linear degradation prediction MSE using large training set, $\lambda = 0.01$.

Dataset	Bearing	TST-S	TST-M	TST-L
FEMTO	Bearing1_3	0.00478	0.00239	0.00213
FEMTO	Bearing1_4	0.01347	0.01523	0.01789
FEMTO	Bearing1_5	0.00236	0.0017	0.00424
FEMTO	Bearing1_6	0.00299	0.00236	0.00179
FEMTO	Bearing1_7	0.00627	0.01264	0.00693
Mean		0.00597	0.00686	0.0066
XJTU-SY	Bearing1_3	0.02925	0.04054	0.05987
XJTU-SY	Bearing1_4	0.21479	0.11055	0.08669
XJTU-SY	Bearing1_5	0.06269	0.03819	0.03099
Mean		0.10224	0.06309	0.05918
XJTU-SY	Bearing2_3	0.00636	0.00885	0.00808
XJTU-SY	Bearing2_4	0.02985	0.027	0.02921
XJTU-SY	Bearing2_5	0.00863	0.01216	0.01645
Mean		0.01495	0.016	0.01791

B. Fault Classification

All of the subsequent investigation were conducted using the TST-M variant depicted in the previous section (Table III). The MLP head was altered to

feature three feedforward layers with rectified linear unit (ReLU) non-linearities and a softmax function following the final feedforward layer. In total, the model variant features 2.6M parameters.

Table VII: Fault classification head.

Layer	Activation Function	Output Dimension
FF1	ReLU	512
FF2	ReLU	128
FF3	Softmax	4

The models were trained with an exponentially decaying learning rate ($\gamma = 0.9$, 1×10^{-4} initialization), batch size 8 and weight decay of $\lambda = 0.001$. Furthermore, an encoder dropout of 0.1 was used. Like the previous experiments, training was halted after three consecutive epochs without improvements in test set accuracy.

Employing a 25% train-test-split, the model was tested in three scenarios: accelerometer data, microphone recording and sensor fusion. The results of these experiments are shown in Table VIII.

Table VIII: MCI dataset fault classification accuracy.

Dataset	Accelerometer in %	Microphone in %	Sensor Fusion in %
LB	99	98.33	98.86
LT	86.58	88.67	90.25
LM	93.42	82.17	96.92

It is evident that classification under laboratory conditions is quite trivial with any of the sensor setups, as each of them performed above 98% accuracy. The other environment scenarios reveal that introducing external disturbances results in a drop in classification accuracy. Most notably, the MCI-LT accuracy drops significantly, with the accelerometer performing worse than the microphone. Surprisingly, the MCI-LM environment, featuring disturbances by two types of rotating machinery (lathe and mill), sees greater classification accuracy in the accelerometer and sensor fusion scenarios. Although microphone performance drops significantly. While this variance in accuracy seems somewhat arbitrary, it is to be expected with real world measurements under a highly dynamic environment, such as a workshop.

Overall, it can be concluded that the model is effective in identifying characteristic signal patterns of different bearing conditions. In fact, TST does so even in a volatile environment using a contact-less, non-intrusive acoustic measurement with a worst-case accuracy of $> 82\%$.

V. CONCLUSION

This thesis proposed and validated TST, a novel transformer-based rolling-element bearing condition modeling framework. The framework was tested in two different use-cases, internal degradation modeling and fault classification. It was shown that degradation modeling performance of TST is adequate on small datasets, but scales well with dataset size. The inclusion of a time token based on the current operating time allowed the model to effectively capture vibration-temporal patterns in the training data, resulting in recurrent neural network (RNN)-like performance without being reliant on continuous monitoring data during inference. Furthermore, TST demonstrated robust fault classification even while severe environmental disturbances were present. These experiments proved that using acoustic emissions to assess the internal condition of a rolling-element bearing is a viable option.

Since the framework has been shown to scale well with dataset size, future work in the field could see to apply TST to even larger datasets. If both quality and quantity of data is ensured, the proposed architecture may perform very well in real, industrial applications.

REFERENCES

- [1] Z. Xu, Y. Guo, and J. H. Saleh, "Remaining useful life prediction with uncertainty quantification: Development of a highly accurate model for rotating machinery," *arXiv.org*, 2021.
- [2] S. Suh, J. Jang, S. Won, M. S. Jha, and Y. O. Lee, "Supervised health stage prediction using convolutional neural networks for bearing wear," *Sensors*, 2020.
- [3] L. Nie, L. Zhang, S. Xu, and W. C. and Haoming Yang, "Remaining useful life prediction for rolling bearings based on similarity feature fusion and convolutional neural network," *Journal of the Brazilian Society of Mechanical Sciences and Engineering*, 2022.

- [4] E. Sejdić, I. Djurović, and J. Jiang, "Time–frequency feature representation using energy concentration: An overview of recent advances," *Digital Signal Processing*, 2009.
- [5] C. Wang, W. Jiang, X. Yang, and S. Zhang, "Rul prediction of rolling bearings based on a dcae and cnn," *Applied Sciences*, 2021.
- [6] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," *arXiv.org*, 2017.
- [7] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv.org*, 2021.
- [8] A. Karpathy, "Connecting images and natural language," Ph.D. dissertation, Stanford University, 2016.
- [9] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016, <http://www.deeplearningbook.org>.
- [10] P. Nectoux, R. Gouriveau, K. Medjaher, E. Ramasso, B. Chebel-Morello, N. Zerhouni, and C. Varnier, "Pronostia : An experimental platform for bearings accelerated degradation tests," *IEEE International Conference on Prognostics and Health Management*, 2012.
- [11] B. Wang, Y. Lei, N. Li, and N. Li, "A hybrid prognostics approach for estimating remaining useful life of rolling element bearings," *IEEE Transactions on Reliability*, 2020.



Hansi Strobel is a Master's student of the Mechatronics program at MCI Innsbruck, Austria. He also worked as a deep learning engineer, which sparked the idea for this thesis. Besides machine learning, he also has an interest in motion control, robotics and embedded software development.
email: hansistrobel@hotmail.de

Immersive co-located visuo-haptic feedback in 3D space

Fabian Triendl, and Yeongmi Kim (supervisor)

Abstract—Most virtual reality feedback devices focus on simple vibration feedback or kinesthetic feedback via a glove or an exoskeleton, but few devices exist that focus on feedback by lateral translocation of the fingertips. In this paper, a novel visuo-haptic feedback device is introduced, which allows for immersive experiences in a virtual reality environment.

The aim is to enable user controlled positioning of a virtual finger in 3D space, while also enabling haptic feedback in the case of a physical collision within the virtual environment.

The haptic feedback provided to the user is composed of two parts, a platform that elevates on vertical collisions and a two axis sled that dislocates the users finger according to the direction of the collision. Communication of the device with the virtual environment is done via a Bluetooth interface and the virtual environment itself is created and implemented in Unity.

The implementation of the device shows a delay between the user indicating movement and the corresponding axis motor starting between 0.04 s and 0.12 s, as well as a controlled axis speed of 15 mm/s for movements along the horizontal plane. The movement range of the device is: X-, Y-axis 37 mm; Z-Axis 12 mm. User testing shows that the device's weight is still within comfortable levels but further improvements can be made in the future.

Index Terms—handheld, haptic devices, immersion, virtual reality, visuo-haptic feedback.

I. INTRODUCTION

VIRTUAL reality is becoming more popular recently and is an application implemented in 3D space where visual feedback is usually provided

Fabian Triendl studies at the Management Center Innsbruck, e-mail: f.triendl@mci4me.at, Yeongmi Kim is with the Department of Mechatronics, MCI, Innsbruck, Austria.

through the use of a headset. The main advantage is that the user is much more immersed in the virtual environment than with the use of traditional screens. This is achieved by the combination of the screens in the headset moving with the head of the user and the camera in the virtual environment moving accordingly, to mimic real life circumstances for the users perception.

Since the user perception not only depends on visual feedback, immersion can be further increased through the implementation of haptic feedback. In this field, small advancement has been made so far, where examples of available solutions are:

- A pen shaped surgical simulator [1]
- A glove pulling the fingers back [2]
- A glove providing vibration feedback [3]
- A glove pulling the fingers back and providing soft touch feedback [4]

It can be seen that available devices focus on simple one directional feedback or vibration for the fingers, but none on lateral translocation of the users finger in three directions. Virtual reality application opportunities can be found within the field of education, where [5] shows that compared to traditional audio-visual media, virtual reality supported learning has a higher individual learning outcome. With virtual reality supported learning, the user attention can be monitored through eye observation and the contents pacing, substantial focus and didactic method can be varied individually in real time to achieve the highest learning result.

Another possibility is the field of tourism, where sites can be virtually visited, without the need of

travel and the danger of tourists damaging historically important heritages. [6] found that user experience under properly adjusted mediation regarding value perception and psychological response showed promising results, even though previous visitation of a site had a negative impact on user perception.

II. METHODS

The creation of the device can be divided into two methodical sections, the design of the hardware and the creation and integration of the software.

A. Hardware Setup

The hardware needs to be able to be mounted onto a virtual reality control device and therefore should be as light as possible. It also has to be able to move a low weighted object, such as a finger, along three orthogonal axes. To ensure high precision and device stability, the device must be able to carry out the motion along these axes with comparably low friction and low play in all directions where it is not desirable.

Size of the final device is also of utter importance to ensure it remains as compact as possible, while still providing a sufficient movement range to ensure a satisfying user experience. To drive the electrical components of the device, a circuit board is created and an external power source is used, which is mounted onto the users arm. As an approach to realize the device, the model of a Cartesian robot is chosen, as can be seen in Fig. 1.

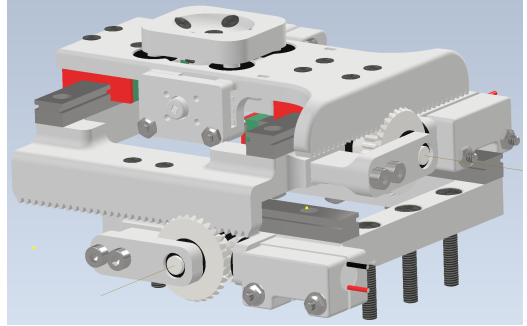


Fig. 1. 3D Design of the Cartesian robot part of the device.

1) *Component Selection:* When selecting the components used for the device, size and weight play a large role on what component is finally chosen. The motors used (EXP-R25-570) are selected with these criteria in mind, and also the required momentum M_m , nominal current I_n and nominal voltage V_n , as well as the rotational output speed v_m . The final motor chosen has the following characteristics:

- Type: Brushed DC, Weight: 1.3 g
- Diameter: 6 mm, Length: 21 mm
- $V_n = 6\text{ V}$, $I_{Stall} = 0.4\text{ A}$
- $M_m = 0.088\text{ Nm}$, $v_m = 90\text{ RPM}$

The linear guide mechanism per axis is chosen to be a linear rail with a roller bearing sled, since this option provides high stability against rotational forces paired with a comparatively low static friction coefficient when compared e.g. with a pure sliding bearing. The used linear guides have the following characteristics:

- Material: Hardened stainless Steel
- Width: 7 mm, Height: 4.7 mm

To control the position of the users finger, the position of each individual axis has to be measured. To achieve this, one of the gears is connected directly to a rotary position sensor for each axis. To keep the size and cost of the final device to a minimum while still maintaining a high precision, a resistance based sensor (Bourns 3382H-1-103) is chosen instead of an encoder. Since such sensors have a certain range

where the internal contact pins are traversing from connected to not connected and hence the output value can fluctuate in these regions, entering these sections is not desirable for this application. This can be mitigated by setting up the gearbox in such a way that the maximum range of motion for each linear axis is within one full rotation minus the undesirable region.

2) *Gear Design*: User testing shows that an adequate linear speed for the X- and Y-Axes v_{XY} is approximately 30 mm/s. To save space, allow maximum movement range in both directions of each linear axis and enable more freedom when designing the linear motion properties of the device, an open gearbox is implemented.

A bevel gear is used, to allow the motor to be placed along the corresponding axis direction, instead of having a 90 degrees offset where the motor is taking up more space. Also it provides a lower friction coefficient compared to e.g. a worm gear, and a lower transmission ratio, which proved to be a better fit for this application. The second gear of the bevel gear is connected to a spur gear, which moves the linear axis through the use of an integrated gear rack. The final gear box can be seen in Fig. 2



Fig. 2. Two part gearbox, the motor is connected to the left bevel gear.

The setup of the gearbox has to follow certain criteria:

- Gear module has to be in a range that allows for stereo-lithography (SLA) 3D printing using tough printer resin (PV-RT-B405-0500-BK) and accounts for play
- Transmission change from 90 RPM to \approx 30 mm/s

- One full rotation of the spur gear covers whole linear motion range
- Diameter of the spur gear small enough to keep the bevel gear transmission within acceptable range

The bevel gear is set up with a chosen module of 0.6. This value is empirically determined by running tests with 3D printed gears and has proven to be a good common ground between small gear size and transmission stability. The transmission of the bevel gear is set to $N = 2:1$. The facewidth of the bevel gear is set to 2.2 mm, providing enough space for the inner spur gear while still large enough to ensure long term stability of the gear. The spur gear module is empirically determined to be $m = 0.55$, to allow operation within the aforementioned criteria. This amounts to a spur gear pitch d_t of

$$p_t = m \cdot \pi = 0.55 \cdot \pi = 1.7279 \text{ mm.} \quad (1)$$

The amount of teeth for the spur gear is set to $n = 22$, which amounts to a movement range d_m per rotation of

$$d_m = p_t \cdot n = 1.7279 \text{ mm} \cdot 22 = 38.01 \text{ mm.} \quad (2)$$

The device is designed in such a way that a movement range for the X- and Y-axes is 36 mm, accounting for the dead space of the rotary position sensor. This results in a movement speed v_{XY} of

$$v_{XY} = \frac{v_m}{2} \cdot d_m = \frac{90 \text{ RPM}}{2} \cdot 38.01 \text{ mm} = 28.51 \text{ mm/s.} \quad (3)$$

The maximum linear force F_l , neglecting friction, provided by the motor can be calculated as

$$F_l = \frac{(M_m \cdot N)}{\frac{d_g}{2}} = \frac{0.088 \text{ Nm} \cdot 2}{0.006 \text{ m}} = 29.33 \text{ N,} \quad (4)$$

where d_g is the outer diameter of the spur gear minus half a tooth height ($t_h/2$) = 0.6 mm) and N is the transmission of the bevel gear. Testing shows

that this is enough force to even move the finger of a user trying to actively resist the movement.

3) *Finger sensor*: To enable user input with one finger, a sensor is implemented to determine if the users finger is pressing more in one direction than in the other. Initially, the sensor concept is chosen between different implementation candidates, which are:

- Resistive
- Capacitive
- Force sensing resistor (FSR) based

For the capacitive approach, the effect of the users finger onto a present electric field is to be measured. Since the effectual movement of the users finger is rather small (below 1 mm) and the influence of parasitic elements, this approach is the hardest to implemented of the three.

Measuring the resistance through the users finger involves a simple and low cost setup, where five screws are integrated into a indentation, as seen in Figure Fig. 3. A voltage is applied to the central pin and the resistance through the users finger is measured for all four input directions. This allows for easy detection of the users finger as well as detection of the pressure applied in each direction, since testing done with a prototype shows that the finger resistance varies between 1 MΩ and 10 MΩ for most people, depending on the applied pressure.



Fig. 3. Pin layout of the resistance based finger sensor.

Measuring the pressure applied by the user’s finger involves a similar layout as the approach where the

finger resistance is measured, but there is no central pin and the outer pins are swapped with FSR (Force Sensing Resistor) sensors (Ohmite FSR04). This can be seen in Figure Fig. 4. Here, the four FSR sensors are placed tilted towards the finger, to allow for easier measurement of horizontally applied forces.

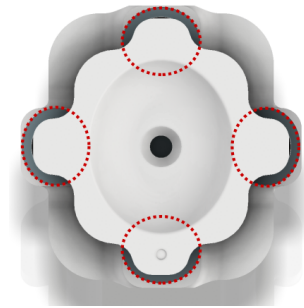


Fig. 4. Layout of the FSR finger sensor.

The FSR sensor approach is chosen in the end since it pairs reliability with compactness and lies within an acceptable price range.

4) *Electrical setup*: The electrical setup of the device consists of a circuit board, made specifically for this device, three motors, one vibration motor, three position sensors, a power supply battery and an *Arduino Nano 33 BLE*, which is used as micro-controller and has integrated *Bluetooth* capabilities. A schematic view of the used setup can be seen in Fig. 5 below.

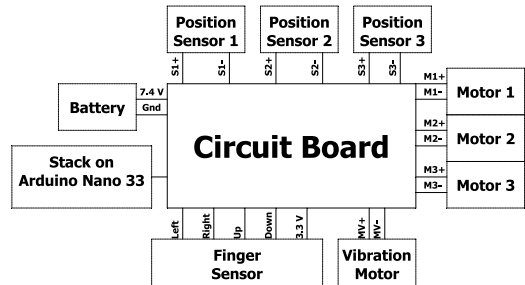


Fig. 5. Electric schematic of the device setup

The circuit board itself is designed with the software *Altium* and has to fulfill the following demands:

- Drivers for three DC motors and one vibration motor
- Circuitry for the three position sensors
- Circuitry for the finger sensor
- Voltage regulator providing 6.5 V for the motors
- Stack on capability for the *ArduinoNano*

Since the *Arduino Nano 33 BLE* only has 5 PWM capable outputs ([7]), the motor drivers have to be able to be controlled with only one PWM pin per motor, which is why a motor driver with a directional pin is chosen. Additionally, the driver also has to fulfill the electrical characteristics of the motor and supply (Stall current $I_{Stall} = 0.4$ A, Motor nominal Voltage $V_m = 6$ V, Signal Voltage $V_s = 3.3$ V and desired control frequency $f_c = 20$ KHz). Due to these factors, the motor controller *ROHM BD63565EFV* is chosen, which can handle two motors each, meaning two controllers are mounted onto the board.

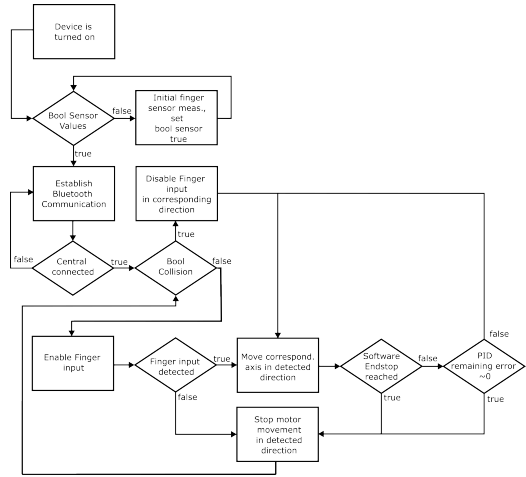


Fig. 6. Flow chart showing the behavior of the device

In order to implement a position controller, as required for the control of the devices end effector on an occurring collision, the system has to be described mathematically. A comparison between the measured and approximated system can be seen in Fig. 7.

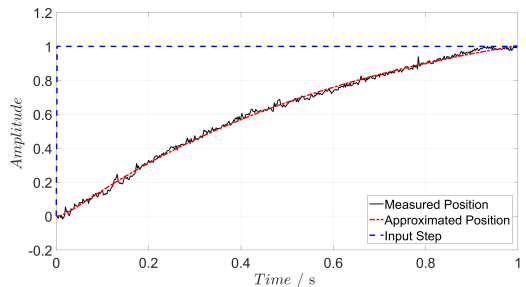


Fig. 7. Comparison between the measured and the approximated position.

B. Software Setup

1) *Microcontroller implementation*: Fig. 6 shows the general behavior of the program written for the function of the microcontroller.

In continuous time, the transfer function of this system is approximated to be

$$T(s) = \frac{1.306}{0.01163s^2 + 0.6866s + 1}. \quad (5)$$

To allow for position control, one PID-controller is implemented for each axis. The PID values are manually fine tuned to achieve an acceptable rise time, settling time and no overshoot. The controlled system response can be seen in Fig. 8

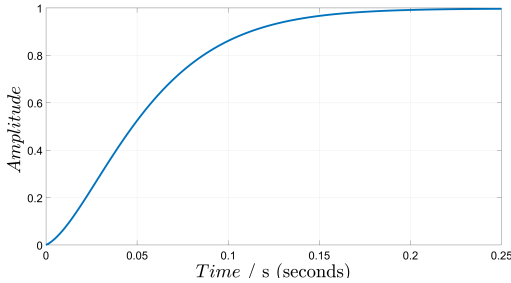


Fig. 8. Controlled response of the simulated system using adjusted PID-values.

2) *Unity implementation:* The virtual environment in unity is composed of three main components. These consist of the Bluetooth communication, the virtual finger following the position of the real one and collision detection of the virtual finger combined with adequate movements. As a prove of concept, a hot wire game is implemented within Unity, as seen in Fig. 9. In this game the user has to move the virtual finger through a set distance between two copper wires. If the users finger touches one of the two wires, the finger is moved away in a realistic way and a short burst of vibration feedback is provided.

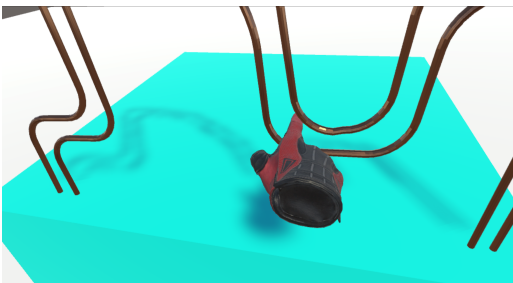


Fig. 9. Picture of the hot-wire game implemented in Unity.

C. Final Setup

The device consists of a base plate, mounted onto a virtual reality controller, the movable x- and y-axes where the x-axis is mounted onto the base plate and the y-axis is mounted onto the x-axis and the finger rest, which is mounted onto the z-axis fixed onto the y-axis. An exploded view of the final assembly can be seen in Fig. 10. The circuit board is mounted onto the back side of the base plate and the battery pack is mounted onto the users hand with a hook and loop fastener.

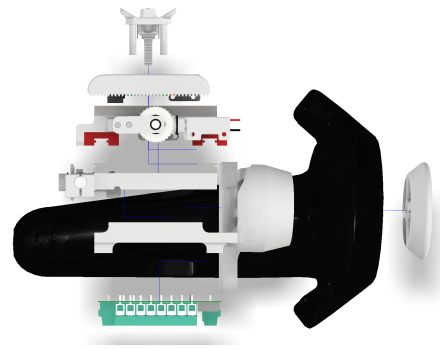


Fig. 10. Exploded view of the device.

The device has the following characteristics:

- X-, Y-axis movement speed: 28.51 mm/s
- X-, Y-axis movement range: 36 mm
- X-, Y-axis Sensor ADC LSB: 0.0186 mm
- Z-axis movement speed: 28.27 mm/s
- Z-axis movement range: 10 mm
- Z-axis Sensor ADC LSB: 0.01 mm

III. EVALUATION

A. Measurements

When testing the final prototype, which can be seen in Fig. 11, the following basic mechanical specifics are found:

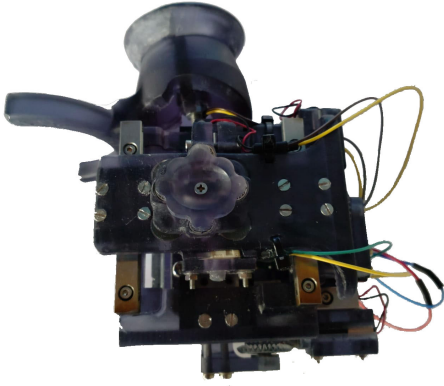


Fig. 11. Final prototype.

- Total weight: 315 g
- Overall dimensions: 120 mm x 84 mm x 74 mm (Width x Height x Depth)
- Maximum movement range: X-, Y-axis 37 mm; Z-Axis 12 mm
- Limited movement range: X-, Y-axis 34 mm; Z-Axis 10 mm
- Movement speed: X-, Y-axis 14 mm/s, Z-axis: 15 mm/s

The controlled response of the X-axis can be seen in Fig. 12. The target position is varied to show the performance of the controlled system during a realistic situation.

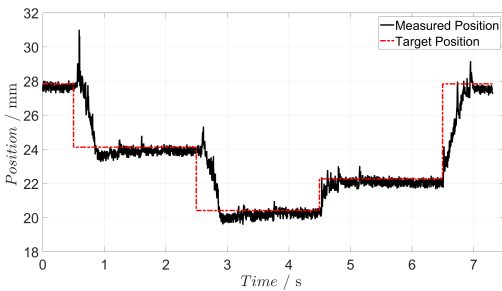


Fig. 12. Controlled response of the real system using the final PID-values.

The input to output delay of the X-axis can be seen in Fig. 13. In this measurement, the axis movement engages only after the finger sensor measurement reaches a value below the set threshold. Once this value is reached, through the user pressing to one side with the finger, the axis begins to move.

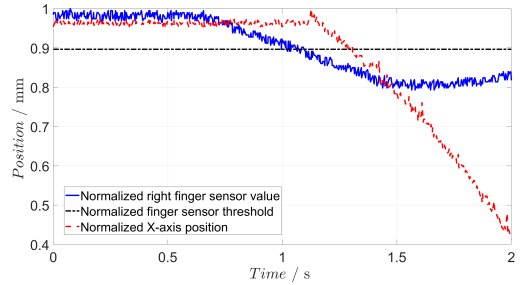


Fig. 13. Finger sensor input compared to the measured axis position.

From this data, an axis reaction speed between 0.12s and 0.04s can be measured. Considering the moment the user input begins, the time between the user finger starting pressing until the motor engages is approximately 0.4s. The reaction time can be improved by reaching the threshold faster, meaning if the user presses in the direction faster than done as shown within this measurement.

B. Comfort

Testing has shown that the user's finger placement requirements often varies from user to user. This implies that there is no one size fits all solution, due to smaller and larger hands or varying user finger flexibility. To alleviate this, an easily adjustable mounting component of the device shall be created in the future. In this approach, two joints are added to the mounting component, which can be easily adjusted to allow for two dimensional rotation of the whole device in relation to the controller it is mounted on.

IV. CONCLUSION

The prototype created fulfills the demands required of it as described in the introduction of this paper. Testing has shown that the movement range of the device is optimal and no requirement exists to enlarge the range as long as the purpose of the device is not altered, e.g. to include another finger or the whole hand. The speed of the devices axes, when controlled by the corresponding PID controller, is satisfactory for the target application.

The finger sensor itself has two approaches implemented. The resistance based approach and the FSR sensor approach. The resistance based approach has the advantage of extremely low cost and mechanical simplicity but had to be abandoned due to the approaches susceptibility to external influences lowering the measurement accuracy.

The software implementation within the game engine Unity is still a work in progress, but will be finished in the very near future. So far the Bluetooth communication and basic finger movement is implemented, which will be further iterated upon and the addition of collision detection will complete this task.

Since a large part of the components of the device are 3D printed, it can easily be manufactured and its overall price is quite low, especially when compared to existing systems. A further decrease to the cost of virtual reality appliances is of utter importance, since the possibilities of virtual reality applications are huge and almost every field can benefit from improvements to the state of the art, be it the field of education, science, hardware/software development, entertainment and many more.

This would allow for the implementation of virtual solutions of real world appliances for a fraction of the cost and a higher operating efficiency. An example would be to familiarize people with a topic in the virtual world, without having to use the otherwise required real world equipment. The possibilities are sheer endless and one can only wait and see what the future brings.

ACKNOWLEDGMENT

The author would like to thank the amazing team of the MCI-Mechatronics department for their now year long support, as well as his girlfriend, for her support and repeated proofreading.

REFERENCES

- [1] (2018) The FundamentalVR website. [Online]. Available: <https://fundamentalsurgery.com>
- [2] (2021) The VRGluV website. [Online]. Available: <https://www.vrgluV.com/enterprise>
- [3] (2021) The ManusVR website. [Online]. Available: <https://www.manus-vr.com/haptic-gloves>
- [4] (2021) The HaptX website. [Online]. Available: <https://haptx.com/virtual-reality/>
- [5] D. A. Muhammad Roy Aziz Haryana, Sony Warsono and E. Nahartyo, "Virtual reality learning media with innovative learning materials to enhance individual learning outcomes based on cognitive load theory," *The International Journal of Management Education*, vol. 20, Indonesia, 2022.
- [6] X. J. a. Xiaojun Fan a and N. Deng, "Immersive technology: A meta-analysis of augmented/virtual reality applications and their impact on tourism experience," *Tourism Management*, vol. 91, China, 2022.
- [7] Arduino S.r.l, "Arduino® Nano 33 BLE Product Reference Manual SKU: ABX00030 (Datasheet)," Italy, 2022.



Fabian Triendl is a student at the MCI, determined to finish his masters degree and develop amazing things in the future.