FACHHOCHSCHULE
WIENER NEUSTADT
University of Applied Sciences – Austria



PROCEEDINGS 2023

Mechatronik | Mechatronics

Kurzfassungen der Masterarbeiten
Master Studiengang Mechatronik
Fachhochschule Wiener Neustadt

Abstracts of Master Theses
Master's programme Mechatronics
University of applied sciences Wiener Neustadt

IMPRESSUM

Proceedings 2023 | Kurzfassungen der Masterarbeiten
Masterstudiengang Mechatronik

Herausgeber: Prof. (FH) Dipl.-Ing. Wolfgang Haindl

Aus Gründen der besseren Lesbarkeit wird auf die gleichzeitige Verwendung männlicher & weiblicher Sprachformen verzichtet. Sämtliche Personenbezeichnungen gelten gleichwohl für beiderlei Geschlecht. Irrtümer & Satzfehler vorbehalten

Copyright FH Wiener Neustadt GmbH, 2023: Alle Rechte, auch das des auszugsweisen Nachdruckes, der auszugsweisen oder vollständigen Wiedergabe, der Speicherung in Datenverarbeitungsanlagen, der Verbreitung, der Bearbeitung & Übersetzung sowie jeder Form von gewerblicher Nutzung, vorbehalten. Alle in diesem Tagungsband enthaltenen Angaben wurden nach bestem Wissen zusammengestellt & mit Sorgfalt kontrolliert. Dennoch sind Fehler nicht ganz auszuschließen. Aus diesem Grund sind die im vorliegenden Tagungsband enthaltenen Informationen mit keiner Verpflichtung oder Garantie irgendeiner Art verbunden. Herausgeber & Institut übernehmen infolgedessen keine Verantwortung & werden keine daraus folgende oder sonstige Haftung übernehmen, die auf irgendeine Art aus der Benutzung dieser Informationen – oder Teilen davon – entstehen, auch nicht für die Verletzung von Patentrechten, die daraus resultieren können.

FH Wiener Neustadt GmbH
Johannes Gutenberg-Str. 3
2700 Wiener Neustadt
Tel.: +43 5 0421 1

Foto-Credits: FH Wiener Neustadt, stock.adobe.com

Satzherstellung: FH Wiener Neustadt, Johannes Gutenberg-Str. 3
2700 Wiener Neustadt

Druck: Products4more Vertriebs- & Systemges.m.b.H.
Molkereistraße 6 2700 Wiener Neustadt

INHALTSVERZEICHNIS

Vorwort 5
Auswertealgorithmik zur Überführung von Prozessfehlern aus Inline-Messdaten beim LSS in ein Volumenmodell – Mohannad Ajamieh 7
Bildauswertung bei hochauflösenden Fahrzeugscheinwerfern – Christoph Babacek 10
Design of a self-tuning regulator for a brushed PMDC Motor – Zoltán Bek BSc 15
Berechnung der Schmierfilmdicke in einem Kegelrollenlager mittels eines datenbasierten Modells – Gigov Boris 24
High-frequency IC design of a command encoder for enhanced particle detection in monolithic active pixel sensors at atlas, CERN – Sebastian Haberl 30
Development of a closed loop hall sensor for high current measurements in MVDC grids – Michael Holzbauer 34
Control electronics for on-chip laser spectroscopy – Lisa Knöbelreiter 40
Entwicklung eines Prüfstandes für Hochleistungsprojektoren – Johannes Kraft 44
Einfluss von Innenstrukturen und Stützstruktur-Parametern auf die Hirtisier®Behandlung von LBM-Bauteilen – Ana Kokic 48
Entwicklung eines Prototypen zum automatisierten Verzwirnen von Wolle für den privaten Gebrauch – Paul Isaak Leitner BSc. 54
Entwicklung intelligenter Erdankersensorik zur Messung des Feuchtegehaltes und der Farbe von Beton – Mustafa Oktay 57
Entwicklung, Simulation und Optimierung eines vierbeinigen Roboters – Sándor Pataky 61
Parksperr Modellierung und Echtzeitsimulation – Philipp Petsch 67
Framework für die Analyse von hochdimensionalen Daten aus dem PBF-LB/M Prozess – Johannes Picker 70
Design of a tunable microwave source for integrated quantum technology – Felix Schinkovits 72
Modellierung der Entstehung und Abtragung von Triboschichten in Reibprozessen – Victoria Seidl 78
A PCB-based approach towards high-speed data transmission and power management for the BCM' system inside the Atlas experiment of the LHC – Anna Swoboda 84
Untersuchung eines Kühlkörpers zur Verbesserung der Kühlleistung – Thomas Thalhammer 88
Analyse eines KI Algorithmus zur Erkennung von Fahrspuren – Marjan Torkamanjahangiri BSc 94
Analyse und Automatisierung eines Hochlasttribometers – Vattappillil Bony 99
Prozessüberwachung im 3D Betondruck: Messung, Aufzeichnung und Verarbeitung von Prozessdaten – Janina Völkl 105

FOREWORD

2024 is a special year for the Austrian University of Applied Sciences sector. It has been 30 years since the first University of Applied Sciences programs started in the fall of 1994. Wiener Neustadt and Mechatronics have been part of it from the very beginning. While Mechatronics was a young and unknown field back then, it has now become one of the key technologies of the 21st century and a central element of the Industry 4.0 vision.

Obviously, we are quite far from a stand still – the networking of machines and organizations, including human involvement, opens up diverse possibilities for innovative solutions in almost all aspects of our lives, such as mobility, health, environment, and energy. Recognizing these opportunities and implementing them responsibly is crucial.

We are therefore pleased to present this collection of outstanding master's theses from the year 2023 in the Mechatronics programme. This volume does not only serve as a representative example of the diversity and width of the field of Mechatronics, but also showcases the abilities of our students to address current technical challenges and develop innovative solutions.

My sincere thanks go to the company mentors and professors who, through their guidance and support, have contributed to the creation of these works. Their professional expertise and support have undoubtedly helped students to tap into their full potential.

VORWORT



Prof. (FH) Dipl.-Ing.
Wolfgang Haindl
Studiengangsleitung

2024 ist ein besonderes Jahr für den österreichischen Fachhochschulsektor. Es ist mittlerweile 30 Jahre her seitdem im Herbst 1994 die ersten Fachhochschulstudiengänge gestartet sind. Wiener Neustadt und auch die Mechatronik waren vom ersten Tag an dabei. War die Mechatronik damals noch eine junge und unbekannte Wissensdisziplin so ist sie heute eine der Schlüsseltechnologien des 21 Jahrhunderts und zentrales Element der Vision Industrie 4.0.

Vom Stillstand sind wir daher weit entfernt – durch die Vernetzung von Maschinen und Organisationen unter Einbeziehung des Menschen ergeben sich wieder vielfältige Möglichkeiten für innovative Lösungen in nahezu allen unseren Lebensbereichen wie z.B. Mobilität, Gesundheit, Umwelt & Energie. Die sich dadurch ergebenden Chancen gilt es zu erkennen aber auch entsprechend verantwortungsvoll umzusetzen.

Sehr stolz präsentieren wir Ihnen diese Sammlung von hervorragenden Masterarbeiten des Jahres 2023 im Studiengang Mechatronik. Dieser Band ist nicht nur ein repräsentatives Beispiel für die Vielfalt und Breite des Fachgebiets Mechatronik, sondern auch für die Fähigkeiten unserer Studierenden, sich den aktuellen technischen Herausforderungen zu stellen und innovative Lösungen dafür zu entwickeln.

Mein herzlicher Dank gebührt natürlich auch den Firmenbetreuern und Professoren, die durch ihre Anleitung und Unterstützung die Entstehung dieser Arbeiten mit ermöglicht haben. Ihre fachliche Expertise und Unterstützung haben den Studierenden sicherlich dabei geholfen, ihr volles Potenzial auszuschöpfen.

AUSWERTEALGORITHMIK ZUR ÜBERFÜHRUNG VON PROZESSFEHLERN AUS INLINE-MESSDATEN BEIM LSS IN EIN VOLUMENMODELL

Ajamieh, M.

Kurzfassung: *Bei der Laserstrahlschmelztechnologie wird ein Laser (Light Amplification by Stimulated Emission of Radiation) als Energiequelle verwendet, um das Metallpulver Schicht für Schicht zu schmelzen und ein Objekt herzustellen. Der Hauptteil des Laserlichts wird vom Metallpulver absorbiert, um es zu schmelzen, ein anderer Teil wird reflektiert. Anhand des reflektierten Lichts könnte der Verlauf des Druckprozesses überwacht werden, um die Fehlstellen zu detektieren. Ein laufendes Projekt der „Forschungs- und Technologietransfer GmbH" (FOTEC) beschäftigt sich mit der Qualifizierung und Qualitätsüberwachung von Laserstrahlschmelzen-Verfahren. Das Qualitätskontrollsystem lieferte verschiedene Parameter, welche durch den Druckprozess kontinuierlich gemessen wurden.*
In dieser Arbeit sollen die Messdaten weiterverarbeitet werden, um relevante Auffälligkeiten zu detektieren und dem Anwender als Volumenbild darzustellen. Die adaptive Schwellwertbildung wurde zur Objektklassifizierung auf die Messdaten angewendet, um die Auffälligkeiten isolieren zu können. Ein Voxelisierungsprozess im Volumenbild wurde durchgeführt, um ein Voxel-Modell der Auffälligkeiten zu erzeugen.
Darauffolgend soll das Voxel-Modell der Auffälligkeiten in einer Feedbackschleife mithilfe eines CT-Bilds verifiziert werden.
Schlüsselwörter: *Laserstrahlschmelzen, Qualitätskontrolle, adaptive Schwellwertbildung, Volumenbild, Voxelisierung.*

1. EINLEITUNG

Für die Qualifizierung und den Einsatz von Laserstrahlschmelz-Verfahren (od. eng. Laser Beam Melting Process) in der Weltraumfahrt, Luftfahrt, Automobilbranche und der Medizintechnik sind Systeme notwendig, welche Informationen über die Qualität der gefertigten Bauteile liefern können. Diese Arbeit wird im Rahmen eines laufenden FFG-Projekts der FOTEC abgehandelt. Das Projekt beschäftigt sich mit der Qualifizierung und Qualitätsüberwachung von LSS-Verfahren.

Die FOTEC hat drei Qualitätskontrollsysteme entwickelt, die in einer LBM-3D-Anlage EOSINT M 280 installiert sind. In der verwendeten LBM-3D-Anlage kommen die Messsysteme, in zwei Kategorien, zum Einsatz:

1. Kategorie 1: Pulverbettüberwachung
2. Kategorie 2: Schmelzbad-Überwachungssystem

Eines der Schmelzbad-Überwachungssysteme umfasst zwei Photodioden, welche das zurückreflektierte Licht detektieren und viele Informationen über den Prozessverlauf liefern. Die Prozesssignale werden aufgezeichnet und stehen somit zur Offline-Auswertung bereit. In dieser Arbeit wurde die adaptive Schwellwertbildung auf die Prozesssignale des Schmelzbad-Überwachungssystems angewendet, mit dem Zweck die auftretenden Bauteilimperfektionen zu isolieren. Die isolierten Imperfektionen wurden mittels eines Voxelisierungs-Verfahrens verarbeitet, um das 3D-Voxelmodell der Auffälligkeiten zu erzeugen. Das 3D-Voxelmodell wurde im STL-Format (Standard Triangle Language) gespeichert. Schlussendlich können in einer Feedbackschleife aus CT-Daten die tatsächlichen Fehlstellen extrahiert, verortet und mit dem gegebenen STL-Modell verifiziert werden.

2. PROBLEM- UND AUFGABENSTELLUNG

Das Schmelzbad-Überwachungssystem umfasst zwei Photodioden, wobei eine Photodiode „on-axis" und die andere „off-axis" verbaut ist (Kern, 2022). Die Photodioden wurden außerhalb des Druckraums montiert und liefern Messdaten für jede Schicht des Druckprozesses. Innerhalb des Druckraums befinden sich zwei Sichtfenster, durch die das reflektierte Prozesslicht vom Druckraum zur Photodiode geführt wird. On- und off-axis Aufnahmen ermöglichen unterschiedliche Beleuchtungswinkel (siehe Abbildung 1).

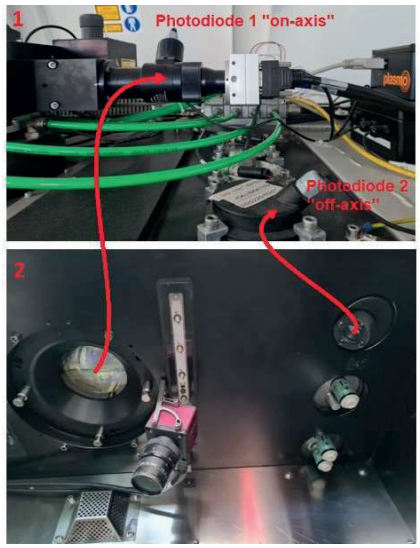


Abbildung 1: In Bild 1 sieht man die zwei Photodionen „on-axis und off-axis". In Bild 2 sind die zwei Sichtfenster innerhalb des Druckraums dargestellt.

Aus den Sensorwerten der on- und off-axis Photodioden werden jeweils die Charakteristiken, Kurzzeitschwankung und Signaldynamik mit unterschiedlichen digitalen Filterkonfigurationen berechnet (FOTEC Forschungs- und Technologietransfer GmbH, 2020). Aus jeder Photodiode werden zwei verschiedene Bilder-Kategorien gespeichert:

- off-axis-Diode: Kategorie 1 und Kategorie 2
- on-axis-Diode: Kategorie 3 und Kategorie 4

Für die Verarbeitung wird nur eine Kategorie von Bildern benötigt. Aus diesem Grund wurden die vier Kategorien

getestet, um diejenige auszuwählen, die bessere Ergebnisse erzielt.

Nach der Auswahl kann der Prozess der Fehlererkennung durchgeführt werden. Um die Defekte im 3D-gedruckten Objekt isolieren zu können, ist es sehr wichtig, eine Methode zu finden, die es ermöglicht die Fehler in den vom Schmelzbad-Überwachungssystem erfassten Bildern zu definieren. Die Messdaten wurden weiterverarbeitet, um relevante Auffälligkeiten zu detektieren und dem Anwender als Volumenbild darzustellen.

Darauffolgend soll das Volumenbild der Auffälligkeiten in einer Feedbackschleife mithilfe eines CT-Bilds verifiziert werden.

3. MATERIALIEN UND METHODEN

Adaptive Schwellwertbildung ist ein Verfahren zur Objektklassifizierung. In diesem Verfahren werden Schwellwerte an die Grauwerthistogramme der Daten angepasst. Nach der Anpassung mit den Grauwerthistogrammen, werden die Grauwerte zwischen benachbarten Schwellwerten zu Aquidensiten zusammengefasst. Wenn die Aquidensiten sich zu bestimmenden Klassen in den Grauwerten des Datensatzes eindeutig unterscheiden, können diese Aquidensiten verschiedene Klassen von Objekten darstellen (Spektrum Akademischer, 2014).

Anhand der adaptiven Schwellwertbildung wurde ein Matlabcode entwickelt, welcher auf die automatisierte Defekterkennung in den Bildern des Überwachungssystems abzielte.

MATLAB abgeleitet von MATrix LABoratory, ist ein Programm, das auf die Verarbeitung der numerischen Daten spezialisiert ist. Die Daten treten z.B. als Messwerte oder bei der Simulation physikalischer Systeme auf. Befehle können direkt im MATLAB Command-Window eingegeben werden. Für unzählige und ausgefallene Berechnungen stehen auch fertige Befehle zur Verfügung. MATLAB bietet eine eigene Programmiersprache, die an gängige Standards angelehnt ist. Damit lassen sich sogenannte MATLAB-Skripte erstellen, um die Funktionalität von MATLAB mit eigenen Befehlen zu erweitern (Angermann, Beuschel, Rau, & Wohlfarth, 2005).

Für diese Arbeit wurde MATLAB Version: 9.9.0.1857802 (R2020b) verwendet, um den benötigten Code zu entwickeln. Der Matlabcode sollte die prozessbedingten Fertigungsfehler identifizieren, isolieren und die so gefundenen Pixelwerte in eine 2D-Matrix eingetragen. Wenn dies für jede Schicht geschieht, erhält man schließlich eine 3D-Matrix, welche die Topologie des Fehlers abbildet. Da es jedoch schwierig ist die Dimensionen des Bauteils bzw. Defekts zu bemessen, wurde die 3D-Matrix in ein Voxelmodell umwandelt. Nach der Umwandlung der Punktwolkendaten in 3D Voxel-Modelle, können die Dimensionen des Objektes bemessen werden, da ein Voxel eine genau definierte Dimension aufweist.

4. PRAKTISCHE DURCHFÜHRUNG

Durch den optischen Vergleich des CT-Bilds in einer bestimmten Region mit einem Bild aus dem Schmelzbad-Überwachungssystem könnten die Bereiche von Defekten in den Bildern des Überwachungssystems definiert werden. (siehe Abbildung 2).

Nach der optischen Erkennung wurde versucht die automatisierte Detektion von Auffälligkeiten durchzuführen. Diese wurde anhand des Grauwerts der Defekte ausgeführt, welche optisch erkannt worden waren.

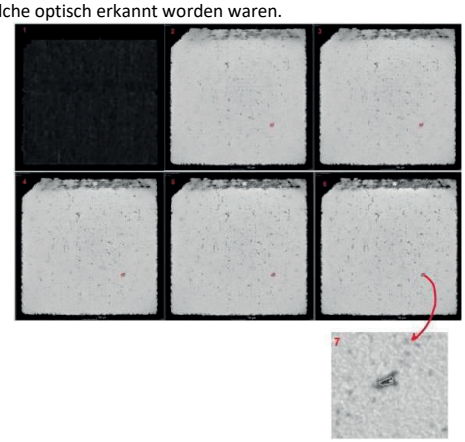


Abbildung 2: Bild 1 ist aus dem Schmelzbad-Überwachungssystem aufgenommen und wurde mit fünf Bildern (2 bis 6) aus dem CT-Bild verglichen.

Der Grauwertbereich muss streng definiert sein, damit ein einheitlicher Bereich verwendet werden kann. Die Berechnung des Grauwertbereichs erfolgte durch die Bearbeitung von 25% der gesamten Bilder der Kategorie 4. Mit Hilfe des Befehls „Thresholding an image“ wurden die minimalen und maximalen Schwellenwerte für 34 Bilder der Kategorie 4 ermittelt.

Nach dem Definieren des Schwellenbereichs konnte die Bildverarbeitung durchgeführt werden. Die Verarbeitung der Bilder des Überwachungssystems zielte darauf ab, die Defekte in jedem Bild zu isolieren und anschließend ein 3D-Modell für alle definierten Defekte zu erstellen. Die Umsetzung gliederte sich in vier Hauptschritte:

1. Der erste Schritt bestand darin, jedes Bild zu verarbeiten, um die Defekte zu isolieren. Die Verarbeitung lief wie folgt ab:

a) Bildreduktion, um nur Bilder für interessante Regionen anzuzeigen, in denen die Probe vorhanden ist.

b) Der zuvor definierte Schwellenwertbereich, wurde auf den ROI angewendet. Dadurch wurden die Defekte im Bild definiert.

c) Um nur die definierte Fehlstelle in den Bildern zu haben, wurde eine Binarisierung durchgeführt. Die restlichen Details wurden gelöscht.

2. Der zweite Schritt ist das Erzeugen eines Volumenbildes aus allen mit dem ersten Schritt verarbeiteten Bildern.

3. Der dritte Schritt ist die Umwandlung des Volumenbildes in ein Punktwolkenmodell. Die Matrix besteht aus binären Daten und durch das Punktwolkenmodell wurde jede Position, an der der Wert eins ist, durch einen Punkt dargestellt.

4. Der vierte Schritt ist die Umwandlung der Punktwolke in ein Voxel-Modell. Um ein 3D-Modell für die Defekte zu erstellen, wurde jeder Punkt aus dem Punktwolkenmodell durch Voxel ersetzt. Die Umwandlung wurde mit der Funktion „VoxelPlotter“ durchgeführt. Abschließend wurde das Voxelmodell anhand der „make_STL_of_Array“ Funktion, im STL-Format gespeichert, damit das Modell per CAD-Anwendung für weitere Messungen verwendet werden kann.

5. ERGEBNISSE

Nachdem der Matlabcode ausgeführt wurde, wurde ein 3D-Voxelmodell erhalten und im STL-Format gespeichert. In Abbildung 3 wurde das STL-Modell dargestellt wobei die dunkelgraue Region dem Grauwertbereich entspricht und die Defektstellen enthält. Nach dem Vergleich, wurde festgestellt, dass die Auffälligkeiten isoliert und in dem 3D-Voxelmodell inkludiert sind. Es ist auch eindeutig, dass viele andere Regionen in diesem Modell integriert sind und die Abweichung enorm ist.

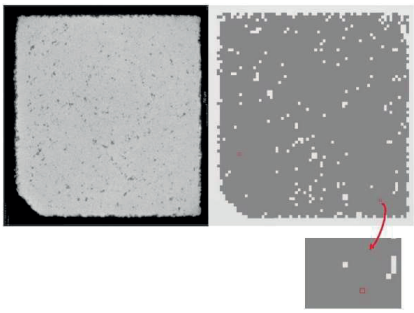


Abbildung 3: Links: CT-Bild; rechts: STL-Modell der Fehlstellen, wobei die rot markierten Regionen die Fehlstellen sind.

Der Grauwertbereich wurde für Defekte in 34 Bildern durchgeführt und für jedes Bild ein eigener Wert berechnet. Leider unterschieden sich die gemessenen 34 Werte voneinander und demzufolge resultierte ein sehr großer allgemeiner Bereich mit ebenfalls großer Abweichung.

6. ZUSAMMENFASSUNG UND AUSBLICK

Das Schwellwertverfahren ist eine zuverlässige Methode für die Detektion der Auffälligkeiten mit dem man eine hohe Genauigkeit erreichen kann, vorausgesetzt der Grauwert des Pixels ist homogen. Der Grauwert des Pixels spiegelt alle Details, die das Pixel enthält, wider. Kleinere Pixel können einzelne Details des Bildes enthalten und den Grauwert dieses spezifischen Details darstellen. Andererseits könnte das größere Pixel viele Details des Bildes enthalten und als Ergebnis wird sein Grauwert in Abhängigkeit von allen enthaltenen Details berechnet (siehe ***Abbildung 4***).

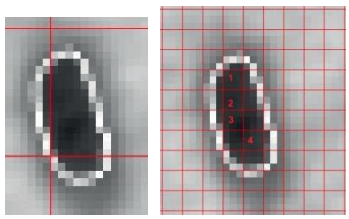


Abbildung 4: Links: vergrößerte Region, welche einen Defekt enthält mit einem 100 µm ×100 µm Raster; rechts: dieselbe Region mit einem 18 µm ×18 µm Raster.

Der Einfluss der Pixelgröße ist äußerst kritisch für den resultierenden Grauwertbereich. Je kleiner die Pixelgröße ist, desto schmaler war der verwendete Grauwertbereich.

Es ist möglich, bei einer höheren Grauwertauflösung, einen schmaleren Schwellwertbereich, und eine kleinere Abweichung für die erkannten Defekte zu erreichen.

7. LITERATURVERZEICHNIS

Angermann, A., Beuschel, M., Rau, M., & Wohlfarth, U. (21. 3 2005). *matlabbuch*. Abgerufen am 10. 9 2022 von http://www.matlabbuch.de/matlab/#matlab

FOTEC Forschungs- und Technologietransfer GmbH. (2020). *FFG-Programm/Instrument: ASAP 13.* UNVERÖFFENTLICHTES DOKUMENT.

Kern, S. (2022). Dimensionsreduktion mehrdimensionaler Daten von digitalen Laser-Beam-Melting-Prozesszwillingen mittels t-Distributed Stochastic Neighbour Embedding. *Bachelorarbeit*. Wiener Neustadt: FHWN.

Spektrum Akademischer. (4. 12 2014). *spektrum.de*. Abgerufen am 10. 8 2022 von https://www.spektrum.de/lexikon/geowissenschaften/adaptive-schwellwertbildung/261



Mohannad Ajamieh

Schulz Straßnitzki Gasse 3/9, 1090, Wien, 06642026009, mohannad.ajamieh@yahoo.com

Mohannad Ajamieh, geboren 1983 in Syrien, studierte Bachelor in Maschinenbau in Aleppo Universität. Danach arbeitete er in verschiedenen Unternehmen als Teamleiter und Berater in den Bereichen Produktion und Instandhaltung. Inzwischen studiert er Masterstudium „Mechatronik“ mit der Spezialisierung „Additive Manufacturing“ an der Fachhochschule Wiener Neustadt.

BILDAUSWERTUNG BEI HOCHAUFLÖSENDEN FAHRZEUGSCHEINWERFERN

Babacek C.

Kurzfassung: *Um mehr Informationen der Fahrzeugumgebung während der Fahrt bei Dunkelheit zu sammeln wird ein - für die Automobilindustrie - neuartiger Ansatz aus Scheinwerfer und Kamera genutzt, um die Umgebung zu erfassen. Diese Methode hat den Vorteil, dass der Fahrzeugscheinwerfer als inverse Kamera dient und zur 3D-Erfassung nur eine einzelne Kamera benötigt wird. Es werden geometrische Modellansätze und künstliche neuronale Netze untersucht, um die zur Erfassung der Umgebung verwendeten projizierten Kreise (Ellipsen) aus den Aufnahmen zu extrahieren. Dabei steht die Berechnungsdauer und Fehleranfälligkeit der gewählten Algorithmen im Blickpunkt. Eine Kombination aus KNN und einem geometrischen Modellansatz liefert schlussendlich einen Ansatz der sowohl robust gegenüber Störgrößen, als auch ressourcenschonend ist. Die Arbeit geht auf Vor- und Nachteile der untersuchten Methoden ein. Das Ergebnis zeigt die Machbarkeit des untersuchten Ansatzes anhand von Laborexperimenten.*

Schlüsselwörter: *DLP-Scheinwerfer, dynamische Umgebungserfassung, Geländetopologie, Künstliche Neuronale Netze.*

1. EINLEITUNG

In den letzten Jahren sind im Automobilbereich Themen wie Computer Vision (CV) und Maschine Learning (ML) immer präsenter [Stiller et al. 2012]. Mit ein Grund ist der technologische Fortschritt der verbauten Microchips. Dieser beeinflusst auch den Markt der Fahrzeugscheinwerfer und deren Ansteuerungen. So wird versucht den Fahrer mit Hilfssystemen zu unterstützen, unter anderem bei der Nachtfahrt. Durch geschickte Auslegung des Scheinwerfers und einer intelligenten Steuerung ist es möglich nur einzelne Bereiche aus dem Fernlichtkegel auszunehmen, ohne dabei gänzlich abzublenden. Wird die Steuerung mit einer Kamera verbunden, können gezielt Gegenverkehr oder vorrausfahrende Fahrzeuge ausgespart werden, die Grundlage des „blendfreien Fernlichtes".

Um diese Ausblendung möglichst gleichmäßig und für den Fahrer weniger erkennbar zu machen, werden immer mehr und immer kleinere Lichtkegel erzeugt. Dabei werden heutzutage Auflösungen, ähnlich zu Videoprojektoren erreicht. Damit können auch Bilder auf der Fahrbahn projiziert werden. Dies schafft die Möglichkeit auch Muster zu erstellen, wie sie bei Streifenprojektionsverfahren verwendet werden. Wird der Gedanke weiterverfolgt lassen sich mit einer Kombination aus Fahrzeugkamera und Scheinwerfer mehr Informationen zur Umgebung sammeln. Dabei ist es wichtig auf einer beschränkten Hardware, mit dementsprechend geringer Rechenleistung, zuverlässig Projektionsmuster auszuwerten. Dies geschieht in der Arbeit mittels drei Methoden. Alle drei haben sie das Ziel projizierte Kreise und Ellipsen zu erkennen und auszuwerten. Im ersten Schritt wird ein einfaches Umgebungsmodell untersucht. Darauffolgend werden unterschiedliche künstliche neuronale Netzte (KNN) trainiert. Als letzte Variante werden die beiden Methoden optimiert und kombiniert.

Grundlage für die Auswertung bildet immer ein System aus Kamera und Projektor (Scheinwerfer), welches zur korrekten Auswertung kalibriert sein muss. Diese Kalibrierung ist zwingend mittels Graustufen-Bilder umzusetzen, da die derzeitigen Scheinwerfer keine Farbprojektionen unterstützen. Die Methode von [Falcao et al. 2008] erfüllt diese Anforderung.

Mit dem kalibrierten System werden Testbilder aufgenommen und mittels der erwähnten Methoden analysiert. Die Kombination aus geometrischen Modellansatz und künstlichem neuronalen Netz liefert eine rasche und zuverlässige Auswertung. Es lassen sich Ellipsen gut analysieren. Das verwendete YOLOv2-Netzwerk liefert eine Positionsschätzung der Ellipsen im Bild und im Anschluss werden mit der Least-Squares Identifikation, die gefundenen Bildausschnitte näher ausgewertet. Die Systemkalibrierung ohne Farbmuster und die ressourcensparende Auswertung zeigen, dass diese Herangehensweise auch für den Automobilbereich geeignet ist.

Im Straßenverkehr ist es mit dieser Methode möglich, mehr Informationen zur Fahrzeugumgebung zu sammeln. [Jung 2007] zeigt, dass schon ein einzelner Lichtstreifen dazu genutzt werden kann, um aus einem zweidimensionalen Bild eine Auswertung mit Tiefeninformationen zu erstellen. Die Autoren nutzen die Auswertung, um Parklücken zu detektieren. Durch eine Projektion von vielen Lichtstreifen/Kreisen/Mustern lassen sich umso mehr Daten sammeln. Denkbar wäre die Fahrbahn detailliert auf Schlaglöcher oder Bodenwellen zu untersuchen. Auch eine Distanzmessung zum vorausfahrenden Fahrzeug wäre möglich. Dazu genügt eine einfache Kamera hinter der Frontscheib. Stereo- oder Trifokalkameras, wie sie von einigen Automobilherstellern verwendet werden, könnten eingespart werden.

2. PROBLEM- UND AUFGABENSTELLUNG

Durch die immer größer werdenden Sicherheitsanforderungen im Fahrzeug, versucht die Automobilindustrie mit immer mehr Technik den Straßenverkehr sicherer zu machen, indem Unfälle gezielt vermieden oder deren Folgewirkungen gemindert werden. Beispielsweise ist der automatische Notruf, der Spurhalte-Assistent oder auch der Geschwindigkeits-Assistent mittlerweile bei Neuentwicklungen ab 2022 Pflicht [EU 2019]. Dabei werden immer mehr Kameras zur Umgebungserfassung verbaut. Ausgewertet werden sie mittels einer Vielzahl an Steuergeräten. Besonders bei Dunkelheit neigen die - oft sehr kleinen Kameras - zum sogenannten Bildrauschen, wodurch Informationen verloren gehen. Ähnliches geschieht auch im menschlichen Auge, bei schlechter Beleuchtung kann das Gehirn deutlich weniger Informationen aus dem Blickfeld gewinnen. Daher unterliefen auch die Fahrzeugscheinwerfer zahlreichen Wandlungen. Ein Moderner Scheinwerfer bietet einen weitaus höheren Funktionsumfang als Abblend- und Fernlicht. Ein intelligentes Steuergerät erkennt automatisch

vorrausfahrende und entgegenkommende Fahrzeuge und blendet diese aus dem Fernlichtkegel aus. Dies geschieht durch gezielte Deaktivierung einzelner LEDs.

Eine Weiterentwicklung dieser, im langläufigen Sprachgebraucht als Matrix-LED bezeichneten Scheinwerfer, stellen sogenannte digital ligth prozessing (DLP)-Scheinwerfer dar. Es handelt sich dabei um einen digital mirror device (DMD)-Chip, welcher aus vielen einzelnen Teilspiegeln besteht, auch Mikrospiegelarray genannt. Diese einzelnen Spiegel lassen sich ansteuern und kippen dabei um ihre eigene Achse. Diese Technologie wird seit vielen Jahren in Kino- und Heimprojektoren bei der Wiedergabe von Filmen und Präsentationen genutzt.

Wird der Chip von einer LED beleuchtet und die Strahlen mittels einer vorliegenden Optik entsprechend gebündelt, kann ein Bild projiziert werden. Das Bild kann dazu verwendet werden, einzelne Verkehrsteilnehmer aus der Fernlichtverteilung auszunehmen (dunkle Bereiche im Bild). Durch die hohe Auflösung von mehr als einer Millionen Bildpunkten, können auch Grafiken auf der Straße dargestellt werden.

Ein weiteres Anwendungsgebiet, liegt in der digitalen Objektvermessung, den sogenannten Streifenprojektionsverfahren. Diese kommen in der Industrie häufig zum Einsatz. Es werden Tiefeninformationen über Objekte mit einer einzelnen oder mehreren Kameras gesammelt. Der große Vorteil liegt darin, dass die Vermessung kontaktlos erfolgt. Es handelt sich dabei um ein aktives Triangulationsverfahren, da mindestens eine strukturierte Lichtquelle zum Einsatz kommt. Dabei werden Muster auf ein Objekt projiziert. Durch die Geometrie des Objektes verformen sich diese Muster aus Sicht der Kamera. Wichtig ist, dass die Kamera einen Offset zum Projektor besitzt. Die Verzerrung des Projektionsmusters lässt sich auswerten und liefert Daten der Objektgeometrie. So ist es möglich mit nur einer einzelnen Kamera ein dreidimensionales Bild eines Objektes zu erhalten.

Wie beschrieben, findet sich in modernen Fahrzeugen ein ähnliches Setup. Der Scheinwerfer mit DLP-Technologie kann als Streifenprojektor fungieren und die Fahrzeugkamera dient als Auswerteeinheit. Werden Kamera und Scheinwerfer synchronisiert, kann das Streifenmuster sehr kurz eingeblendet werden. Durch die Trägheit des Auges nimmt der Fahrer das Muster nicht wahr. Die Kamera ist meist zwischen Frontscheibe und Rückspiegel verbaut. Somit besteht auch der geforderte Abstand (Offset) zwischen Scheinwerfer und Kamera. Es stellt sich die Frage, wie ein auf der Straße projiziertes Muster im Fahrzeug ausgewertet werden kann. Diese Masterarbeit soll hierzu eine Grundlage bilden.

3. MATERIALIEN UND METHODEN

3.1 Kamerakalibrierung

Grundlage für eine Auswertung projizierter Streifen bildet ein kalibriertes System aus Projektor und Kamera. Das Kalibrieren der Kamera gestaltet sich recht einfach. Hierzu existiert eine Vielzahl an Programmen. Da im Zuge dieser Arbeit das MatrixLaboratory (MatLab) verwendet wurde, findet sich mit der von [Boguet 2003] entwickelten MatLab Toolbox eine passende Software. Als Testaufbau wurde eine Logitech C270 HD Webcam kalibriert. Der Vorteil dieser Webcam liegt neben dem günstigen Preis darin, dass sie einen fixen Fokus besitzt. Würde die Kamera eine Autofokus beinhalten, müsste sie bei jeder Fokusänderung neu kalibriert werden. Es werden neun Bilder eines Schachbrettmusters an unterschiedlichen Positionen aufgenommen und anschließend deren Eckpunkte ausgewählt, wie dies im linken Teil der Abb. 1 zu sehen ist. Im rechten Teil findet sich das Ergebnis der Kalibrierung. Hier werden die neun aufgenommenen Schachbrettmuster dargestellt.

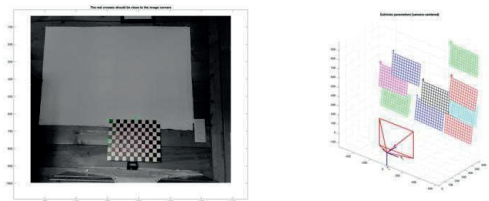


Abb. 1: Ergebnis der Kamerakalibrierung.

3.2 Projektorkalibrierung

Auch der Projektor muss kalibriert werden. Hierzu gibt es ebenfalls bereits erhältliche Software. Oft kommen diese Produkte bei der Vermessung von Bauteilen in der Industrie zum Einsatz. Dabei werden meist Farbmuster verwendet, um den Projektor zu kalibrieren. In dieser Arbeit kommt die Toolbox von [Falcao et al. 2008] zum Einsatz, welche es ermöglicht auch ohne Farbmuster zu kalibrieren. Dies ist notwendig, da die am Markt befindlichen DLP-Scheinwerfer keine Farbprojektion unterstützen. Als Testaufbau wird ein Epson H561 B Projektor verwendet. An diesem wird die bereits erwähnt Logitech Kamera angebracht. In Abb. 2 ist der Testaufbau dargestellt. Der Offset zwischen Kamera und Projektor wurde mit ein PVC-Rohr, welches direkt am Projektor verschraubt wurde hergestellt. Der rechte Teil des Bildes zeigt eine Visualisierung des Ergebnisses in MatLab. Auch hier ist wieder der Abstand zwischen den beiden Systemkomponenten gut zu sehen.

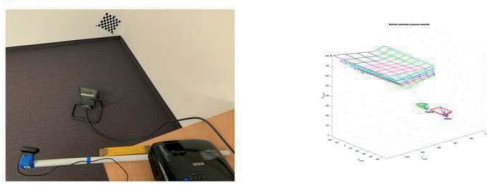


Abb. 2: Aufbau der Projektorkalibrierung im linken- und dem Ergebnis im rechten Bild.

Die Kalibrierung des Projektors läuft ähnlich wie bei der Kamerakalibrierung ab. Zusätzlich wird ein Bild des Schachbrettmusters vom Projektor an die Wand geworfen. Die Kamera nimmt das Muster der Kamerakalibrierung und das projizierte Muster gleichzeitig auf. Anschließend werden in beiden Mustern die Eckpunkte extrahiert.

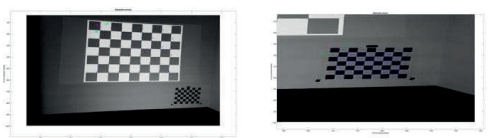


Abb. 3: Ungenaue Auswertung der Eckpunkte im Schachbrettmuster im linken Teil des Bildes und eine Vergrößerung der Auswertung des Kalibrierungsmusters im rechten Bild.

Dabei ist immer die genaue Auswertung der Eckpunkte zu beachten. Werden diese nicht genau genug getroffen, bricht die

Kalibrierung ab. Ein Beispiel für eine schlechte Auswertung findet sich in Abb. 3.

3.3 Auswertungsmethoden

Sind sowohl Kamera als auch Projektor kalibriert, wird eine Auswertung zur Erkennung von Kreisen gesucht. Das klassische Streifenmuster wird vermieden, da in der Praxis viele streifenähnliche Geometrien auf den Straßen existieren (Zebrastreifen, Pfeile, Schriften etc.). Es kommt bei der Projektion von Kreisen jedoch zu einer Verzerrung und die Auswertung muss dementsprechend auch Ellipsen identifizieren. Durch die Vielzahl an verbauten Komponenten, werden die Platzverhältnisse im Fahrzeug immer beengter. Dadurch müssen neue Steuergeräte sehr kompakt dimensioniert werden. Prozessoren mit hoher Rechenleistung schließen sich somit aus. Es muss in der Auswertung dementsprechend darauf geachtet werden, dass möglichst wenig Rechenaufwand benötigt wird.

Mittels der Hough-Transformation (HT) lassen sich Geraden in Bildern sehr leicht finden. [Somonovsky 2021] zeigt, dass sich die HT auch bei der Ellipsenidentifkation nutzen lässt. Grundsätzlich steigt der Aufwand bei einer Analyse von Ellipsen mittels HT zu einem Akkumulatorraum der fünften Dimension. Eine Auswertung benötigt bei hochauflösenden Bildern, selbst unter Verwendung leistungsfähiger Hardware sehr lange. [Somonovsky 2021] zeigt, dass durch einige Annahmen der Raum auf ein eindimensionales Akkumulatorarray verringert werden kann.

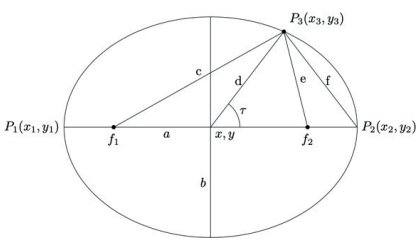


Abb. 4: Beispiel einer Ellipse mit Mittelpunkt (x, y) und Halbachsen a und b.

Betrachtet man Abb. 4, so lässt sich für das Pixelpaar $\boldsymbol{P_1}(x_1, y_1), \boldsymbol{P_2}(x_2, y_2)$ annehmen, dass es auf einer Hauptachse liegt. Somit können sowohl x und y, als auch a und τ berechnet werden. Die Berechnung erfolgt unter der Bedingung, dass die Strecke $\overline{\boldsymbol{P_1P_2}}$ größer als ein bestimmter vordefinierter Wert ist. Anschließend wird ein weiteres Pixel $\boldsymbol{P_3}(x_3, y_3)$ zur Berechnung der halben Länge der Nebenachse herangezogen. Stellt man nun die Ellipsengleichung

$$b^2 = \frac{a^2 d^2 \sin^2(\tau)}{a^2 - d^2 \cos^2(\tau)} \quad (1)$$

in Polarkoordinaten dar, kann die Gleichung nach der gesuchten Länge b der Nebenachse umgestellt werden. Diese Nabenachse kann nun durch ein eindimensionales Akkumulatorarray ermittelt werden.

Eine weitere geometrische Modellbildung, um Ellipsen zu identifizieren findet sich mit der Least Squares Methode (LSM). Sie dient zur Lösung eines überbestimmten Gleichungssystems. Ellipsen zu finden ist ein nichtlineares Problem. Verlangt man aber, dass die Parameter $\boldsymbol{w}$ linear bleiben, können auch nichtlineare Optimierungsaufgaben linear gelöst werden. Die typische Darstellung der LSM lässt sich durch

$$\boldsymbol{y} - \boldsymbol{X}\boldsymbol{w} = 0 \quad (2)$$

Definieren, wobei der Ausgangsvektor $\boldsymbol{y}$ und die Datenmatrix $\boldsymbol{X}$ bekannt sind. Durch Lösen des Gleichungssystems kann der gesuchte Parametervektor $\boldsymbol{w}$ gefunden werden. Es handelt sich dabei um die sogenannte direkte Methode.

Die gezeigten Verfahren gehören zur Kategorie der White-Box-Modelle. In den letzten Jahren steigt das Interesse, wohl durch die steigende Rechenleistung von Computern, an Black-Box-Modellen immer mehr. Dabei sind vor allem künstliche neuronale Netze gemeint (KNN). Es werden in der Arbeit drei unterschiedliche Netze untersucht. Diese unterscheiden sich deutlich voneinander.

Der Aggregate Channel Feature (ACF) benutzt Wavelets um bestimmte Merkmale in Bildern zu extrahieren. Das Eingangsbild wird in verschiedene Einzelbilder aufgeteilt und mit diversen Filtern verändert. In jedem dieser Kanäle wird mittels Fensterschiebeverfahren nach Merkmalen gesucht. Aus den gefundenen Merkmalen, welche mittels AdaBoost-Algorithmus gefunden werden, wird eine Merkmalspyramide erstellt.

Das Region Based Convolutional Neural Network (R-CNN) besteht aus zwei Stufen. Mittels Fensterschiebeverfahren wird grob nach Merkmalen der Trainingsdaten gesucht und daraus eine Heatmap erstellt. Im Anschluss werden alle Regionen mit einem CNN nach Merkmalen gesucht. Dadurch werden auch kleinere Objekte im Bild gefunden. Dieses Netz wird in MatLab als vortrainiertes Netzwerk angeboten. Somit kann mittels Transfer-Learning das Netz weitertrainiert, anstatt neutrainiert werden. Das hat den Vorteil, dass deutlich weniger Trainingsdaten benötigt werden und das Training kürzer dauert.

Mit dem You Only Look Once Netzwerk (YOLOv2) wird ein modernes Netzwerk untersucht, welches ebenfalls vortrainiert geladen werden kann. Ein weiterer Vorteil ist, dass dieses Netzwerk sowohl Klassifizierungs- als auch Detektierungsdatenbanken zum Training verwendet. Auch die Auflösung der Bilder beeinflusst das Verhalten des Netzes nur gering. Mittels Anchor Boxen wird das Bild in einem Schritt ausgewertet. Im einem Klassifizierungsnetzwerk werden zuerst Objekte im Bild erkannt und anschließend mit einem direkt angeschlossenen zweiten Netzwerk auf deren Koordinaten untersucht. Dieses Netz ist deutlich schneller als das R-CNN.

Durch einen Zusammenschluss von einem YOLOv2-Netzwerk und einer geometrischen LSM-Identifikation wird versucht das Beste aus beiden Welten zu vereinen. Einerseits untersucht das CNN das Bild sehr schnell und grob nach Ellipsen ab, um andererseits direkt im Anschluss die gefundenen Ellipsen genauer zu untersuchen. Durch weitere Codeoptimierungen wie das Löschen von Bildpunkten ohne relevante Daten und das Verringern der Auflösung kann die Berechnung weiter beschleunigt werden. Durch die Bewegung des Fahrzeuges, wird auch die Bildgröße weiter verringert, indem weniger Kreise projiziert werden (statt vier Reihen nur eine Reihe).

4. PRAKTISCHE DURCHFÜHRUNG

4.1 1D-Ellipsenidentifikation

Die 1D-Ellipsenidetifaktion mittels HT, ist ohne weitere Adaptierungen sehr anfällig auf Störgrößen. Um dies zu prüfen, findet sich in Abb. 5 eine Testaufnahme. Dabei wird mit dem Projektor ein 8×4 Schachbrettmuster projiziert. Weiters befindet sich ein Schachbrettmuster im Blickfeld der Kamera.

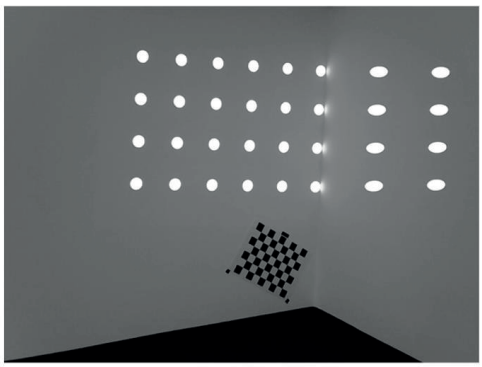


Abb. 5: Eingangsbild der Ellipsenidentifikation.

In Abb. 6 findet man das Ausgangsbild. Deutlich zu erkennen ist, dass die Auswertung vor allem in Bereich des Schachbrettes fälschlicherweise Ellipsen findet. Dieses Verhalten kann durch geeignete Bildbereinigung verbessert werden.

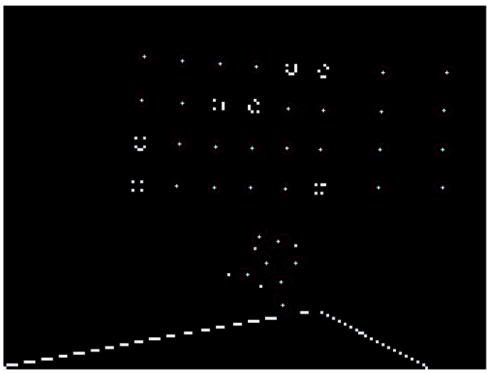


Abb. 6: Ergebnis der 1D-Ellipsenidentifikation ohne Anpassungen.

Durch eine weitere HT, können Linien im Bild gefunden und beseitigt werden, wie in Abb. 7 gezeigt wird. Als Abschluss wird ein morphologischer Filter angewandt, um die letzten kleineren Störungen zu beseitigen.

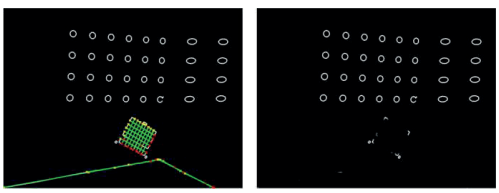


Abb. 7: Ergebnis der HT-Liniendetektion (links) und der anschließenden Bereinigung des Bildes (rechts).

Das Ergebnis in Abb. 8 zeigt, dass der entwickelte Algorithmus alle 32 Ellipsen fehlerfrei findet. Einzig die Orientierung der Ellipsen ist nicht exakt. Die Auswertedauer für das gezeigte Bild beträgt rund 0.76 s.

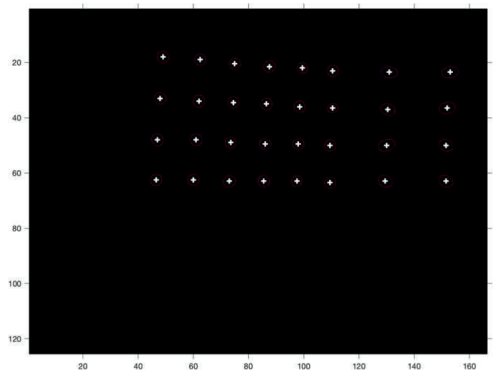


Abb. 8: Ergebnis der 1D-Ellipsenidentifikation mit Bildbereinigung.

4.2 KNN am Beispiel YOLOv2

Als besondere Schwierigkeit stellt sich die Trainingsdatengenerierung heraus. Es finden sich keine dedizierten Datenbanken für projizierte Kreise und Ellipsen. Aus diesem Grund werden die Trainingsdatensätze mit einem MatLab-Code erstellt. Die Ellipsen unterscheiden sich in ihrer Intensität, Größe, Form und Position. Als Auflösung wurden 425 x 319 px gewählt, da das vortrainierte ResNet18 diese Auflösung verwendet und somit der Transfer-Learning-Ansatz verwendet werden kann. Eine Trainingsdauer von nur drei Epochen bei einer sehr geringen Lernrate von $\alpha = 10^{-5}$ stellt sich als optimal für die Aufgabe heraus.

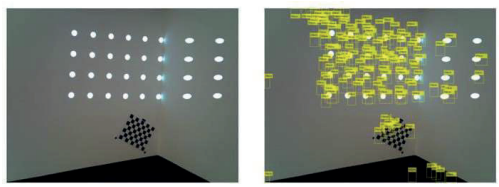


Abb. 9: Eingangsbild (links) und Ergebnis des YOLOv2 mit einem Trainingsdatensatz von 300 Bildern (rechts).

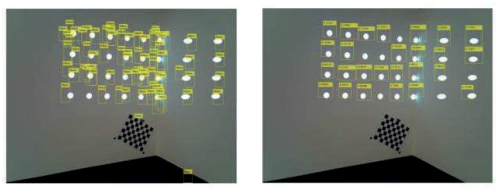


Abb. 10: Ergebnis des YOLOv2 mit einem Trainingsdatensatz von 3000 Bildern (links) und 30000 Bildern (rechts).

In Abb. 9 zeigen sich die starken Unterschiede des Netzverhaltens aufgrund unterschiedlich großer Trainingsdatensätze. Das linke Bild zeigt das Ausgangsbild während auf der rechten Seite die Detektion bei 300 Trainingsbildern zeigt. In Abb. 10 findet sich das Ergebnis mit 3000 Bildern links und 30000 Bildern rechts. Dabei werden 27 von 32 Ellipsen korrekt gefunden. Die Auswertedauer für das gesamte Bild beträgt 0.0323 s. Somit ist das YOLOv2 deutlich schneller als der HT-Algorithmus.

4.3 Kombination aus CNN und geometrischen Modellansatz mit Codeoptimierung

Durch die bereits erwähnte Codeoptimierung wird nur eine Reihe aus Kreisen/Ellipsen analysiert. Die Ergebnisse könnten in der Praxis zwischengespeichert werden. Durch die Bewegung des Fahrzeuges würde sich dennoch eine Karte der Umgebung erstellen lassen. Zusätzlich wird die Ellipsenidentifikation mittels Least Squares Lösung ermittelt und dabei die Bildmatrix weiter verkleinert, nachdem mit dem YOLOv2 Netzwerk das Bild zuerst grob analysiert wird. Im Anschluss werden nur noch die Bildausschnitte mit Ellipsen an den LSM weitergegeben. Die Bildmatrizen werden durch sogenannte Sparse-Matrizen ersetzt.

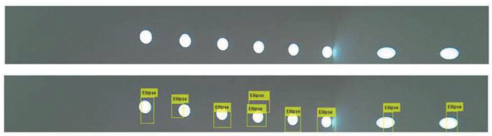


Abb. 11: Eingangsbild (oben) und Ergebnis des YOLOv2 Netzwerkes (unten).

Abb. 11 zeigt das Eingangsbild und das Ergebnis der YOLOv2-Auswertung. Die Ellipsen werden ausgeschnitten, wie dies in Abb. 12 gezeigt wird und an die LSM übergeben.

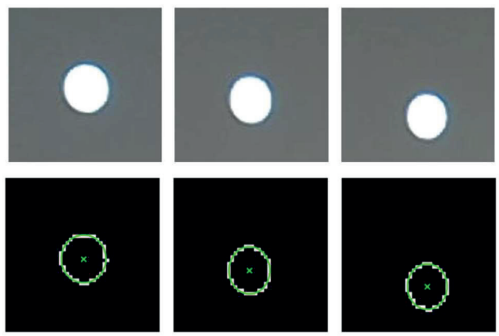


Abb. 12: Bildausschnitte, welche vom YOLOv2 gefunden wurden (oben) und Ergebnisse des Kombinierten Ansatzes (unten).

Der LSM ermittelt im Anschluss sehr exakt die gesuchten Ellipsenparameter. In Abb. 12 werden die Ellipsen und deren Mittelpunkte grün eingezeichnet. Bemerkenswert ist die sehr gute Annäherung an die tatsächlichen Messwerte.

Durch die Codeoptimierung, die Kombination aus KNN und geometrischen Modellansatz und zusätzlich einer Reduktion der Auflösung, erfolgt die Auswertung in 0.01 s.

5. BEWERTUNG DER ERGEBNISSE

In der Arbeit erfolgt eine Bewertung der Algorithmen. Zur Bewertung wurden vier Kriterien herangezogen:

- Auswertedauer
- Genauigkeit
- Umsetzungsaufwand
- Robustheit

Zusammengefasst lässt sich feststellen, dass der reine Blackbox-Modellansatz mittels KNN, die geringste Ausführungszeit aufweist (0.03 s). Ein Nachteil liegt beim Umsetzungsaufwand, da zuerst ein passendes KNN gefunden und trainiert werden muss und anschließend noch zusätzlich eine LSM durchgeführt wird. Die Genauigkeit ist im Vergleich zu den anderen Algorithmen am geringsten. Dieser Nachteil kann mit der Kombination aus geometrischem Modell und KNN behoben werden. Wird zusätzlich eine Codeoptimierung durchgeführt erreicht man eine Ausführungszeit von 0.01 s bei sehr hoher Genauigkeit. Durch die zweistufige Analyse ist dieses Verfahren auch sehr robust.

6. ZUSAMMENFASSUNG UND AUSBLICK

Die gefundenen Verfahren zeigen die Machbarkeit der Idee, einen Scheinwerfer mit einer Kamera zur 3D-Vermessung der Fahrzeugumgebung zu verwenden. Es finden sich zahlreiche Vorteile dieser Technologie. Beispielsweise kann die Anzahl der Kameras im Fahrzeug reduziert werden. Auch können die Messwerte den verbauten Assistenzsystemen wie Abstandstempomat, Parksensoren oder dem Notbremsassistent zur Verfügung gestellt werden, um so die Auswertung zu verbessern. Durch die sogenannte Sensor-Fusion würden sich die Daten von Radarsensoren und Kamera zusammenfassen lassen, um so ein noch detaillierteres Modell der Fahrzeugumgebung zu schaffen.

7. LITERATURVERZEICHNIS

Stiller, C.; Bachmann, A.; Duchow, C.; 2012. „Maschinelles Sehen“. In: Handbuch Fahrerassistenzsysteme: Grundlagen, Komponenten und Systeme für aktive Sicherheit und Komfort. Hrsg. von Hermann Winner, Stephan Hakuli und Gabriele Wolf. Wiesbaden: Vieweg+Teubner Verlag, 2012, S. 198– 222. isbn: 978-3-8348-8619-4. doi: 10.1007/978-3-8348-8619-4_16. url: https://doi.org/10.1007/978-3-8348-8619-4_16

Gabriel Falcao, Natalia Hurtos und Joan Massich. „Plane-based calibration of a projector-camera system“. In: VIBOT Master 9 (Jan. 2008)

H. G. Jung u.a. „Light Stripe Projection based Parking Space Detection for Intelligent Parking Assist System“. In: *2007 IEEE Intelligent Vehicles Symposium*. 2007, S. 962–968. doi: 10.1109/IVS.2007.4290241

Amtsblatt der Europäischen Union. Verordnung (EU) 2019/2144. https://eur-lex.europa.eu/legal-content/DE/TXT/PDF/?uri=CELEX:3201 9R2144&from=HR. 2019.

Jean-Yves Bouguet - Intel Corp. Camera Calibration Toolbox for Matlab. 12. Apr. 2003. url: http://robots.stanford.edu/cs223b04/JeanYvesCalib/.

Martin Simonovsky. *Ellipse Detection Using 1D Hough Transform*. Version 1.2. 11. Okt. 2021. url: https://www.mathworks.com/matlabcentral/fileexchange/33970-ellipse-detection-using-1d-hough-transform



Christoph Babacek

ZKW/FH Wiener Neustadt, Forstgasse 17 2632 Wimpassing, +43 664 882 646 43, christoph.babacek@fhwn.ac.at

Geboren am 10.04.1996 in Neunkirchen, aufgewachsen in Wimpassing im Schwarzatale, BG/RG Sachsenbrunn, HTL Mödling, Bachelor Mechatronik (FH Wiener Neustadt), Software-Entwickler ZKW-Elektronik

DESIGN OF A SELF-TUNING REGULATOR FOR A BRUSHED PMDC MOTOR

Zoltán Bek BSc

Abstract: *The present thesis introduces the basic concepts of adaptive control and examines the advantages of using one of its specific variants called Self-Tuning Regulator (STR) to control the speed of a laboratory setup based on a brushed PMDC motor. The introduction of adaptive control and the STR design method is done by referencing relevant literature, followed by several computer simulations and laboratory experiments with the aim to achieve an efficient and precise speed control resistant against disturbances in all working points. The study shows that the steady state gain of the laboratory setup is dependent on its operating point, furthermore it has mildly non-linear characteristics, its speed is saturated and its parameters change with time. Furthermore, it is stated that the STR controller designed to the target system is able to control the system in all operating points with the desired dynamics, compensate load disturbances and identify parameter changes, however, the determination of the observer polynomial is crucial from the point of view of stability. Finally, further research potentials are identified.*

Keywords: *Adaptive Control, Self-Tuning Regulator, Brushed PMDC Motor, System Identification*

1. INTRODUCTION

The increasing complexity of challenging scientific and industrial problems requires more flexible and more intelligent solutions. Physical system with strongly non-linear characteristics challenge classical control design methods, which often only work in a pre-defined operating point where the system is linearized. Another challenging property of several physical systems is that they have time dependent variable parameters, which means that the original properties of the controlled system can change significantly during its operation resulting in an inaccurate or even unstable behaviour [1]. Motivated by those problems adaptive control has opened a new direction in process control that is significantly different from classical static control.

This thesis is intended to introduce the basic concepts of adaptive control and to demonstrate the advantageous effects of using a specific variant of it called Self-Tuning Regulator (STR). That will be done by providing the detailed controller design steps and the results of different computer simulations and laboratory experiments based on an existing Permanent Magnet DC (PMDC) Motor laboratory setup at the Fachhochschule Wiener Neustadt (FHWN).

2. SELFT-TUNING REGULATOR

Although in the past decade there was a significant growth on the market of the brushless PMDC motors, which offer a much more reliable, low-maintenance alternative to brushed DC motors, there is still a wide application range for PMDC commutator motors. They commonly find application in the automotive branch as auxiliary motors, in the industry as linear actuators or torque motors, in the toy production or even in underwater vehicles [2]. Furthermore, the chosen laboratory setup has an operating point dependent steady state gain with nonlinear distribution over its complete operating range. Moreover, the dynamic behaviour of the laboratory system also varies slightly with the change of the operating point. These properties make the laboratory system a perfect subject for the implementation of an adaptive control.

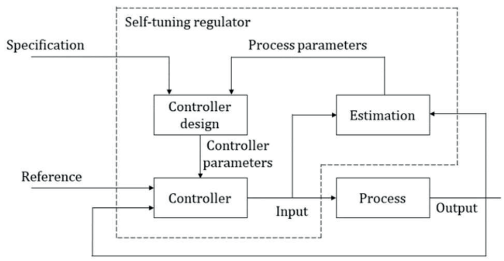


Fig. 1: STR block diagram.

As Fig.1 shows, one of the most striking features of the STR method is that it consists of well-separated units. The *Estimation* block identifies the process parameters and forwards them to the *Controller design* block which calculates the new controller parameters. Those will be forwarded to the *Controller*, which adapts its behaviour and outputs a new control value fitting to the all-time system properties and the required closed-loop system response. From the previously introduced structure it follows that during the design of an STR controller three engineering problems have to be solved. Firstly, for the automated system identification a proper recursive form has to be found which is able to describe the behaviour of the system by a well-chosen mathematical model and to update it permanently. Secondly, a stable and reliable controller design algorithm has to be implemented that is able to calculate the new controller values in real-time. Finally, the controller has to be implemented in a such flexible way that its parameters can be modified during operation and that the accidental changes in its control value do not lead to unstable or undesired closed-loop behaviour.

3. MODEL AND EXPERIMENTAL SETUP

The experimental setup is based on the commercial product *DR300 Laboratory Setup* of the company *amira GmbH* which is primarily designed for educational purposes. A detailed description of the setup is given in its reference manual[3]. As Fig.2 shows the *DR300* consists of a 19" module box (Bx) which serves as an actuator and a mechanical setup including two PMDC motors with one tacho generator and one incremental encoder (E) mounted on the motor shaft of the left motor (M1). The motors are mounted facing each other and their drive shafts are rigidly coupled via metal diaphragm clutch. The right motor(M2) serves as the controllable load of the left one.

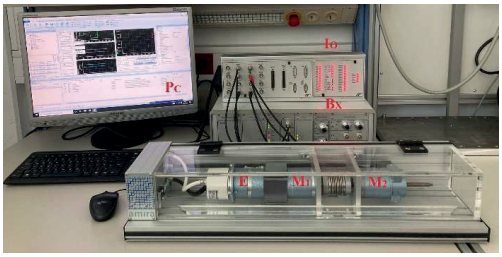


Fig. 2: Laboratory setup.

The module box contains servo amplifiers, built-in current regulators for both motors, a signal adaption unit, power supplies and a module with measurement outputs. The box is extended by an analog PI-controller which is integrated in the case. The setpoint of the current controllers is adjustable by providing DC voltage to their inputs in the range of ±10[V]. To enable a flexible control, the measurement outputs and the current regulator setpoint channels are connected via an IO Box (Io) to an external real-time controller board called *dSPACE DS1104 R&D*. This real-time hardware based on the PowerPC technology is plugged directly into one of the free PCI/PCIe slots of the laboratory computer (Pc). Thanks to the Real- Time Interface (RTI) functionality provided by the company *dSPACE*, controller models created in *Simulink©* can be compiled into target specific C-Code and loaded into the program memory of the real-time hardware. The RTI also enables real-time interaction with the running code via the software *ControlDesk V6*. That way the laboratory setup enables to fully concentrate on the regulator design without having to worry about the peculiarities of the low-level implementation. The following image provides a comprehensive picture about the entire system.

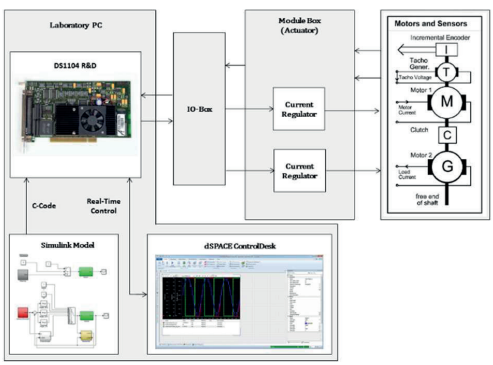


Fig. 3: Laboratory setup schematic diagram.

Assuming that the model structure of a process is known, a Self-Tuning Regulator is able to estimate the process parameters in real time and to design a regulator based on the recently estimated process parameters to obtain the desired closed-loop behaviour [1]. To identify the discrete process parameters in real-time the least squares method was used. In a recursive formulation this identification can be performed permanently when new measurements are available. The main idea is that the parameters of a linear system can be estimated by requiring that the square of the output estimation error converges to minimum [4]. If the input vector *u* and the output vector *y* of an unknown systems with an assumed linear structure are known, then the estimated output vector $\hat{y}$ can be expressed by $\hat{y} = Xw$ where w is the unknown parameter vector and X is the matrix of the independent input data vectors, also called regression matrix or data matrix. In that case the estimation error can be expressed as

$$e_i = y_i - \hat{y}_i = y_i - x_i^{nT} w. \tag{1}$$

Requiring that the square $I(w)$ of the above presented error function converges to its minimum renders

$$I(w) = \sum {e_i}^2 = \frac{1}{2} w^T H w - h^T w + \frac{1}{2} y^T y \rightarrow min. \tag{2}$$

Solving this problem by setting the gradient of $I(w)$ equal to zero renders the well-known solution

$$w = H^{-1} h = (X^T X)^{-1} X^T y, \tag{3}$$

where the expression $(X^T X)^{-1} X^T$ is also known as pseudo inverse. In this thesis the recursive solution of this least squares method was used which is able to estimate the parameters in real time, by updating the estimates in every iteration cycle based on the new measurements extended with a recursive forgetting to exponentially reduce the influence of the old measurement data [4].

To perform an accurate identification, it is essential to have a mathematical model which properly describes the behaviour of the target system. That is why this thesis used the well-known transfer function of the DC motor as a starting point

$$\frac{w}{U_A} = \frac{1}{k_M \phi} \frac{1}{1 + sT_M + s^2 T_A T_M}, \tag{4}$$

where w denotes the rotational velocity of the motor shaft, U_A marks the input (armature) voltage, $k_M \phi$ stands for a machine specific constant, T_M and T_A are the mechanical and electrical time constants of the motor and s is the complex Laplace variable.

Extending the previously presented formula with the documented electrical and mechanical properties of the laboratory setup[3] the general transfer function of the target system is obtained

$$G(s) = \frac{n}{I_{Asp}} \frac{V_{sys}(n)}{(1 + sT_i)(1 + sT_{Mech}(n))(1 + sT_{Meas})}, \tag{5}$$

where n denotes the speed of the motor, I_{Asp} denotes the set point current of the integrated current controllers, $V_{sys}(n)$ stands for the system gain which is speed dependent, T_i and T_{Mech} are the time constants of the electrical and mechanical components of the complete systems and T_{Meas} denotes the damping effect of the sensor system.

Since the STR concept operates in the discrete-time domain the definition of a discrete-time transfer function is necessary. By discretizing the transfer function of the complete laboratory setup the general model structure

$$G(z) = \frac{b_0 z^2 + b_1 z + b_2}{a_0 z^3 + a_1 z^2 + a_2 z + a_3}, a_0 = 1 \tag{6}$$

is obtained. By including some noise transferred by the same system dynamics, the so-called ARX model structure is obtained, which is used during the whole thesis to describe the target system

$$A(z) y(k) = B(z) u(k) + e(k). \tag{7}$$

A direct approach of the STR method to control design is to define the desired closed loop characteristics of the controlled system extended by the definition of the desired response to control signals. Based on these requirements a controller can be determined that creates a closed loop systems together with the process[1]. A simple closed-loop block diagram of a control system can be given by:

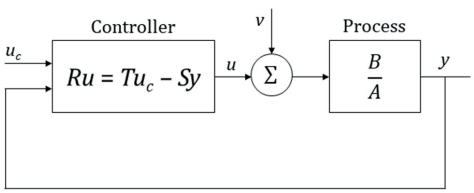


Fig. 4: General linear controller.

After rearranging the equations the closed loop characteristics can be received:

$$y(t) = \frac{BT}{AR + BS}u_c(t) + \frac{BR}{AR + BS}v(t). \tag{8}$$

The numerator forms the characteristic polynomial A_c

$$AR + BS = A_c. \tag{9}$$

During the regulator design this characteristic polynomial can be specified according to the specific system requirements [1]. This gives a huge freedom in design since the characteristic polynomial contains the poles of the closed loop system and the dynamic behaviour of a system is fully specified by the location of its poles [5]. In other words, the system dynamics can be directly specified during the regulator design. Since A and B are estimated permanently after the specification of characteristic polynomial A_c it will be possible to retrieve the polynomials R and S by solving the above equation, which is the so-called *Diophantine equation*, a well-known algebraic problem. It always has a solution if A and B do not have any common factors [1]. Since the *Diophantine equation* determines only R and S the specification of some other conditions is necessary to obtain the missing regulator polynomial T. One suggested method is to define how the desired closed-loop system shall react on the command signal u_c. This can be described by the model following equation

$$A_m y_m(t) = B_m u_c(t). \tag{10}$$

Based on the previous equation a model-following condition can be obtained

$$\frac{BT}{AR + BS} = \frac{BT}{A_c} = \frac{B_m}{A_m}. \tag{11}$$

Leaving the detailed derivation of the STR theory to the literature [1], the control design can be described with the following steps:

1. Defining the dynamics of the desired system by pole placement $A_c = A_0 A_m$ (has to be done only once during regulator design).
2. Ensuring, that the static gain of the reference model will be unity: $B_m\ (1)/A_m\ (1) = 1$. The polynomial B_m shall be chosen accordingly (has to be done only once during regulator design).
3. Estimation of the process parameters B and A.
4. Solving the *Diophantine equation* to obtain *S* and *R*.
5. Calculation of β to ensure steady state gain. $\beta = A_m\ (1)/B(1)$, or alternatively: $\beta = B_m/B$.
6. Determination of $T = \beta A_0$.
7. Computing the control signal u: $Ru(t) = T\,u_c(t) - Sy(t)$

A_0 models the so-called parasitic modes that result from the fact, that the physical behaviour of the real process can not be modelled perfectly. Because of this A_0 is often handled in literature as an additional low-pass effect for reduction of noise characteristics which is not considered by the process model [5]. Since the difference between the process model and the real physical behaviour of the process is use case dependent, A_0 has to be tuned for each use case individually.

A critical point of the real-time controller design is the time-efficient solution of the *Diophantine equation* to obtain the controller parameters R and S. By taking into consideration the chosen general model structure describing the entire laboratory setup and the design rules presented above a general design equation becomes,

$$(z^3 + a_1 z^2 + a_2 z + a_3)(z^2 + r_1 z + r_2) + (b_0 z^2 + b_1 z + b_2)(s_0 z^2 + s_1 z + s_2) = (z^3 + a_{m1} z^2 + a_{m2} z + a_{m3})(z^2 + a_{o1} z + a_{o2})\,. \tag{12}$$

A comparison of coefficients for z renders six equations. By grouping the controller parameters into a column vector, the following system of equations can be obtained

$$\begin{bmatrix} b_0 & 0 & 0 & 1 & 0 \\ b_1 & b_0 & 0 & a_1 & 1 \\ b_2 & b_1 & b_0 & a_2 & a_1 \\ 0 & b_2 & b_1 & a_3 & a_2 \\ 0 & 0 & b_2 & 0 & a_3 \end{bmatrix} \begin{bmatrix} s_0 \\ s_1 \\ s_2 \\ r_1 \\ r_2 \end{bmatrix} = \begin{bmatrix} a_{o1} + a_{m1} - a_1 \\ a_{o2} + a_{o1}a_{m1} + a_{m2} - a_2 \\ a_{o2}a_{m1} + a_{o1}a_{m2} + a_{m3} - a_3 \\ a_{o2}a_{m2} + a_{o1}a_{m3} \\ a_{o2}a_{m3} \end{bmatrix}. \tag{13}$$

The matrix on the left side is known as *Sylvester matrix* in applied mathematics. Based on its definition the *Sylvester matrix* is non-singular only if A and B do not have common factors which is defined as a prerequisite to obtain a unique solution of the *Diophantine equation* [1]. Since the elements of the so-called *design vector* on the right side of (13) contain the pre-defined dynamic characteristics of the closed-loop system and the elements of the *Sylvester matrix* will be obtained as a result of the identification process the above linear equation can be solved by a simple Gaussian elimination [5]. The structure of this design problem remains unchanged for arbitrary model orders which means that building the *Sylvester matrix* and the *design vector* always leads to a straightforward solution as long as the non-singularity condition of the *Sylvester matrix* is fulfilled.

4. PRACTICAL IMPLEMENTATION

The practical implementation can be divided into three sections, identification, computer simulations, and real-time laboratory experiments.

4.1 Identification

The step responses of the open-loop system are measured in the laboratory with several different input voltages. The lowest input voltage is found experimentally by increasing the input in small steps until the motor starts to rotate. The purpose is to cover the whole practical operating range (1000-4000[RPM]). The motor shaft starts to rotate only if the input

voltage is higher than 1.4[V]. However, if it is already rotating the input voltage can be decreased down to 1.1[V]. This indicates the significant role of friction. Furthermore, an input voltage higher than 2.4[V] leads to a saturation. The step response measurements are imported into the *System Identification Toolbox ™* of *MATLAB*. By using the polynomial model estimator functionality the parameters of a chosen model structure can be estimated with the help of the LS estimation method. Based on above the structure *ARX331* is chosen. Comparing the simulated model outputs with the measured step responses it turns out that the *ARX331* structure is able to describe the target systems behaviour in the whole operating range with an average model response fit of 96, 5%, see Fig. 5.

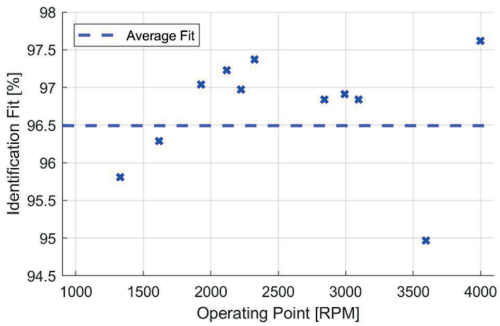


Fig. 5: Model output fits in the identified operating points.

After executing the identification with each data set twelve different transfer functions are obtained. The difference between them gets clearly visible by plotting their unit step responses and bode diagrams, see Fig. 6 and Fig. 7.

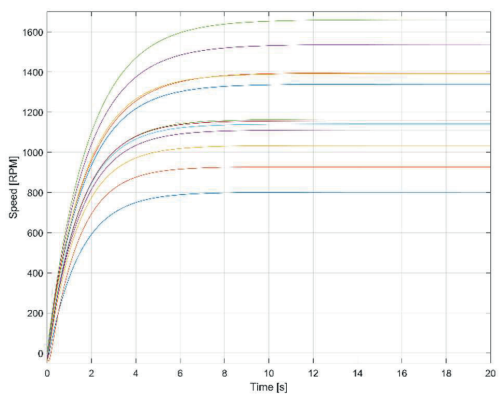


Fig. 6: Unit step responses.

The transfer functions show very similar dynamic characteristics particularly in the low frequency regions. The only noticeable difference is their steady state gain. This means that the input signal amplification of the system depends on its operating point. This recognition fits to the properties of the laboratory setup specified by the manufacturer [3], i.e. that the speed-dependent load resulting from the sliding friction is not negligible. When plotting the unit step steady state values of the transfer functions in relation of the operating points in which they were measured a nonlinear curve is obtained, see Fig. 8. From the results, it can be concluded that the operating point-dependent and mildly non-linear characteristics of the system justifies the need for a controller that identifies the properties of the system at the respective operating point and designs a controller fitting to it. Furthermore, it is also proved that the model structure *ARX331* is suitable for the description of the system behaviour in each relevant operating point.

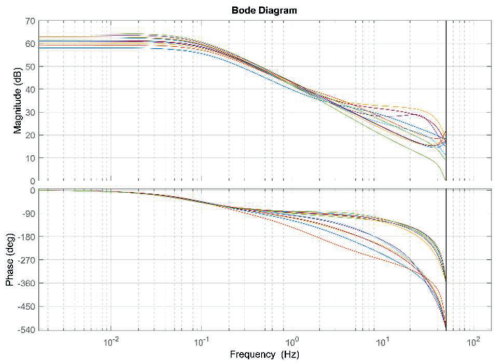


Fig. 7: Bode plots.

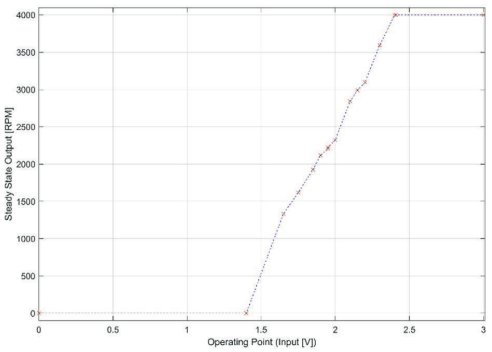


Fig. 8: Nonlinear characteristics of the system.

4.2 Simulations

Due to the complexity of the system identification and the controller design procedures it is desirable to perform simulations before the target system is put into operation. The sampling line $Ts = 0.01[s]$ is chosen as sample time for the simulation and laboratory experiments. Based on the results of the identification chapter the target system is modeled by

$$(z) = \frac{3.921z^2 + 2.022z - 3.806}{z^3 - 1.321z^2 - 0.06804z + 0.3908}. \quad (14)$$

To objectively judge the efficiency of the adaptive controller, a reference PI controller to which the behaviour of the controller can be compared designed via the *Control Systems Toolbox* of *MATLAB* is used. The reference controlled can be represented in parallel-form by

$$K(z) = K_p + K_i + \frac{zT_{samp}}{z-1}, \quad (15)$$

with the coefficients $Kp = 0.00553$ and $Ki = 0.0178$. The expression $(z\,T_{samp})/(z-1)$ represents the *Backward Euler* integration method which has the advantage that it always yields a stable discrete-time system [5]. After designing the reference controller, a similar static design process is executed

to prove that the STR concept is able to produce a similar closed-loop characteristics as the reference controller. Based on the steps described above first the desired closed loop model $(B_m(z), A_m(z))$ is defined with trial and error as long as a closed-loop model can be obtained,

$$G(z) = \frac{0.00116z^2 + 0.0005981z - 0.001126}{z^3 - 2.646z^2 + 2.31z - 0.6635}, \qquad (16)$$

which has similar system dynamics and settling time than the closed-loop system built by the process model and the reference controller. In simulations a dead-beat observer $A_0 = z^{deg\,A0}$ is used as the observer polynomial. Since the polynomials B_m and A_m fulfill the design conditions, and every necessary element is known the *Diophantine equation* can be set up. Solving the equation, building the closed-loop system and eliminating the pole-zero pairs renders a third-order transfer function which is perfectly equivalent to the model transfer function above, which means that the STR design procedure applied on the static model of the target system worked perfectly and created with the help of the controller polynomials a closed-loop system, which fulfills the pre-defined model-following conditions. The bode diagram and the pole-zero map of the obtained closed-loop system indicates that the system is stable since all poles are inside of the unit circle with a 105° phase margin. Furthermore, it is damping higher frequencies even better than the closed-loop system produced with the reference controller but at the same time it has approximately the same settling time. Therefore, the presented design methods are considered legitimate.

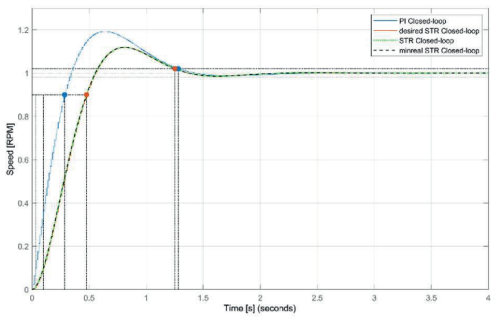


Fig. 9: Closed-loop response of the PI and the STR system.

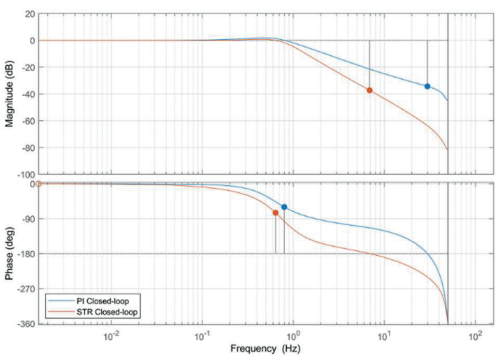


Fig. 10: Closed-loop response of the PI and the STR system.

For further simulations Simulink model is created, using signal-flow diagrams, see Fig. 11. To remain in the discrete-time domain a fixed-step discrete-time solver is applied with a sampling time of $Ts = 0.01[s]$. The most important parts of the *Simulink* model are:

0. Target Value - freely configurable target speed
1. Observer Polynomial Variable - allows the dynamic control of the observer coefficients
2. Target Value Change Detection - resets the identification in case of target value change
3. System Identification - includes the implementation of the RLS identification
4. STR controller design - includes the controller design algorithm
5. STR controller - includes the actual implementation of the controller
6. Controlled Motor - discrete time transfer function of the laboratory setup (process)
7. Load Disturbance - freely configurable motor load

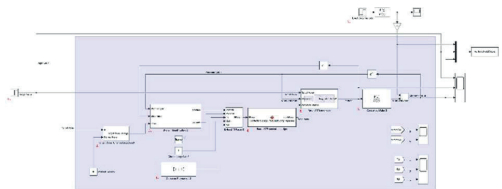


Fig. 11: STR System Simulink Model

One of the most critical parts of the model design is the implementation of the controller, because it has to be flexible to allow future extensions but should avoid the usage of algebraic loops, therefore the implementation in Fig. 12 is used:

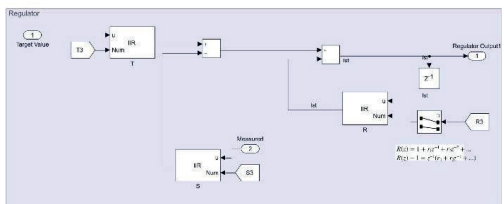


Fig. 12: Implementation of a dynamic STR controller

Load disturbance and actuator saturation tests suggest that the addition of integral action and windup protection is necessary to compensate load disturbances and to avoid fatal integrator saturations, see Fig. 13.

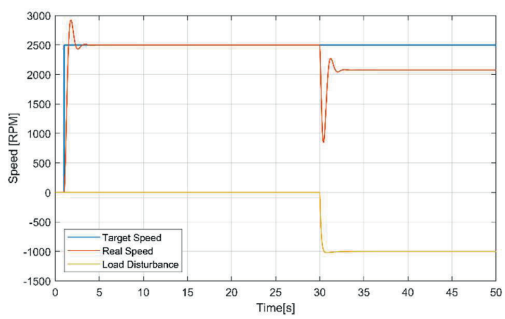


Fig. 13: STR system with uncompensated load disturbance

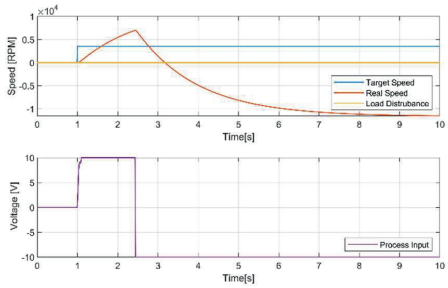


Fig. 14: Integral windup of an STR system with integral action.

The load compensation is realized by introducing an extra closed-loop pole having the effect of integration. Since an integral action has the disadvantage that it can bring the actuators into saturation, the physical limits of the actuators had to be considered in the controller model.

To test the robustness of the newly developed system a test sequence is defined with multiple target value and load steps. Simulation of the test sequence proves that the controller designed with integral action and windup protection is accurate, sufficiently fast and resistant against load disturbance and controller windup, see Fig 15.

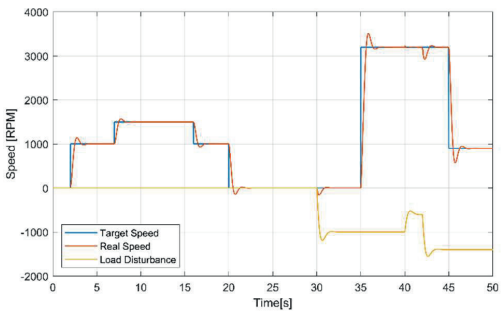


Fig. 15: Test sequence executed on the extended STR system.

Having designed a proper controller, the next step is to extend the simulation with operating point dependent variable process parameters to simulate the behavior of the real target system as realistically as possible. For that reason, the transfer function block simulating the process behaviour is modified to enable the external definition and modification of the numerator and denominator coefficients. The dynamic definition of those coefficients is done by the subsystem *Select Parameter* which switches between the transfer function parameters obtained by the system identification depending on the requested target speed. The parameters of a transfer function are used as long as the requested target speed is smaller than the unit step steady state speed of that transfer function. That way an operating point dependent discontinuous transfer function is obtained, which serves as an acceptable approximation of the real system behaviour.

With the appearance of the variable process parameters two model blocks, *Estimator Anti Windup* and *Target Value Change Detection* become important already included in the original model. These two blocks monitor the changes in the target speed signal and reset the identification if a significant change happens. This so called covariance reset is suggested [1] to achieve a fast parameter convergence and to eliminate the influence of the previously identified values. If the identification is successful and the system achieves the desired target speed the Estimator Anti Windup disables the parameter identification until a new target speed change happens. This is essential to prevent *identification windup*, which could result from inadequate excitation. An important result of the simulation is that the parameter identification converges fast and accurately. However, at the moment of each target speed change the system output shows a strange behaviour before responding with the desired system dynamics. The reason for that is assumed to be the non-minimum phase characteristics of the system transfer functions identified during the system identification. A closer look at the identified step responses reveals, that the target system seems to have a slight delay in nearly each operating point. The *ARX331* system identification delivers a very good fitting in the overall system dynamics and in the steady state behaviour, however in some cases the initial delay becomes even more significant. As mentioned in literature [6] the STR algorithm is sensible against non-minimum phase systems and such a behaviour can lead to control problems. Assuming that the delays resulted only from some inaccuracy in the measurements the transfer function models are replaced with similar system dynamics and steady state gain, however without delays in the system. That way much better results are obtained, see Fig. 17:

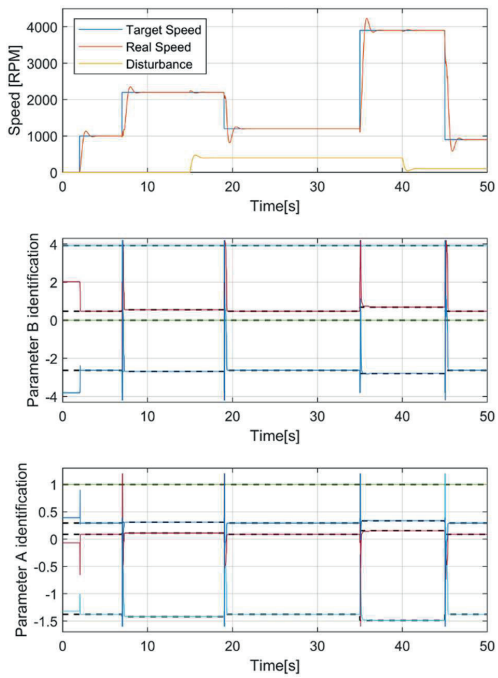


Fig. 16: Simulation of a variable parameter machine with minimum phase transfer functions

The identification remains fast and accurate and the undesirable output jumps disappear. There are still some regulation noises at the beginning of the operating point changes, however that is assumed to be the result of the very fast and discontinuous parameter switching and is considered as acceptable.

4.3. Laboratory Experiments

After proving the legitimacy of the STR design the final step is to test the STR using the laboratory setup. Choosing the proper observer polynomial plays a crucial role in the design of an indirect STR controller [1], since this polynomial is responsible for modelling the parasitic modes and filtering noise and load disturbances. The initial experiments revealed that dealing with load disturbances does not cause significant difficulties if an integral action is included in the controller design. In contrast, the designed control system is particularly sensitive to measurement noise.

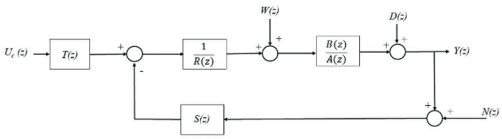


Fig. 17: Modelling of measurement noise

Therefore, noise is in the focus during the observer polynomial design, assuming that a properly chosen observer reduces the noise sensitivity. Assuming that the noise is added at the system output the noise transfer function can be calculated for different observer coefficients, which reveals why the proper choice of the observer coefficients is crucial. Depending on the coefficients the noise damping can be influenced.

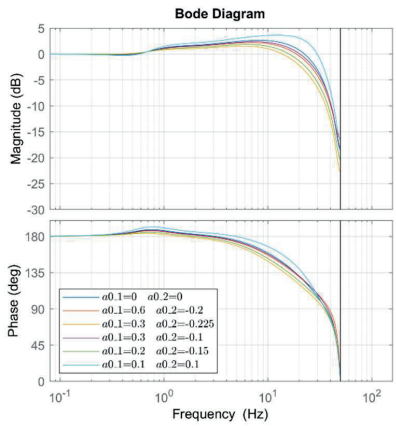


Fig. 18: Bode plot of the different noise transfer functions.

For example using a deat-beat observer $a01 = 0\ a02 = 0$ amplifies the noise much stronger than other observers. Therefore, the observer has to be chosen according to noise damping requirements of the system. The coefficients $a01 = 0.3\ a02 = -0.225$ are identified as a good initial choice.

The most important step in the laboratory experiments is to define test sequences which are able to excite the system such a way that it can be determined if the closed-loop system is accurate in steady-state, provides the required system dynamics, is resistant against load disturbances and works reliable in the whole operating range. Hence, an operating point variation test is performed in which different target value and load steps are executed with different timing. In addition to the first test, the time dependent change of the system parameters can also represent a serious motivation for designing an adaptive controller. Hence, a so-called long run test is executed to determine if the system parameters change in case of a long-term operation without switching the working point.

The operating point variation test consist of four test sequences. In the first test sequence a regular PI controller is used for the startup. It has the purpose to excite the system for initial identification, until the STR controller can take over. This is used as an alternative to the initial parameter set definition suggested by the literature[1]. In the first sequence after the STR regulator has taken over the control only small variations in the setpoint are executed to test if the controller can manage to switch between regular operating points. In the second sequence the system is operated at it limits of 700[RPM] and 3900[RPM], to prove that the control is working properly at the limits of the system as well. In the third sequence relatively small and fast target speed changes are executed in the middle of the operating range to prove that the controller can handle dynamic changes as well. In the last section the resistance of the system against load disturbance is tested by varying the input voltage of the second PMDC motor thereby creating several positive and negative load step situations. The test sequence executions are controlled manually in the GUI of the experiment software *ControlDesk V6*.

At the beginning the sequence fails several times. Those failures are attributed to a low value of the forgetting factor. Using a relatively small forgetting factor worked well during the simulations resulting in a fast parameter convergence, but the simulations did not consider the noise characteristics of the target system. A small forgetting factor results in identifying the noise which disturbs the identification and causes a very intense change in the controller values and consequently renders an unstable behaviour. Since the estimator is reset after every significant target value change and the estimation is only enabled as long as the output value reaches the target value it is not necessary to use a low forgetting factor, because the parameters of the previous estimation are discarded immediately, after the target value change. Therefore, a significantly higher forgetting factor $\lambda = 0.99$ is enough to reach acceptable parameter convergence. An other significant factor is the proper choice of the observer polynomial coefficients. The values $a1 = 0.3,\ a2 = -0.225$ serve as a good starting point since they produce stable step responses. However, sometimes the response becomes almost unstable before automatically stabilizing again.

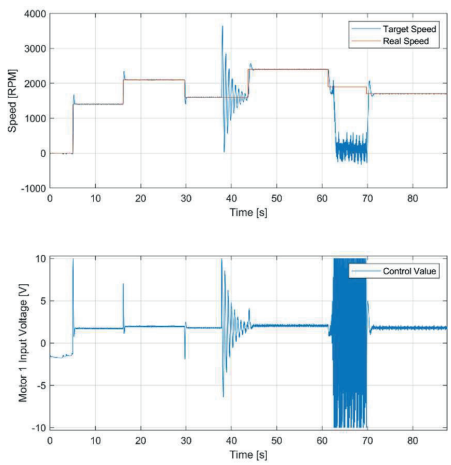


Fig. 19: Operating point variation with badly chosen observer polynomial.

This undesired behaviour is due to badly chosen coefficients of the observer polynomial causing an unsatisfactory noise damping and provoking aggressive controller responses, which lead to unstable behaviour without any change in the parameter values. Since there are no suggestions for finding the proper observer coefficients in literature[1] the trial and error principle is used. That is realized by using the sliders in *ControlDesk* to finetune the coefficient. That way the new observer coefficients $a1 = 0.2,\ a2 = -0.15$ for a stable behaviour are identified. It is clearly visible that the controller is able to successfully drive the system to follow the target value and manage to compensate the disturbances resulting from the uncertainties of the system and load change.

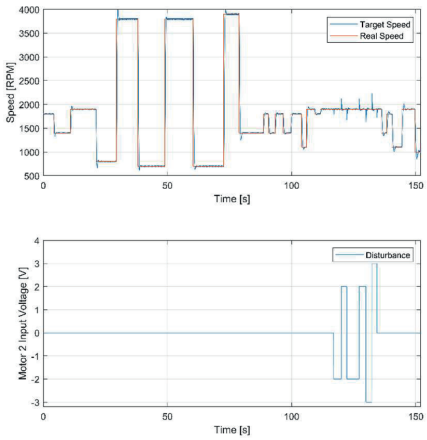


Fig. 20: Operating point variation with well chosen observer polynomials.

Comparing the step of STR and PI controller (designed for an unfortunately chosen working point), enlightens that both controller perform well. Although, up to 1800 – 1900[RPM] both can provide the desired system dynamics with the slight overshoot, on the border of the operating range only the STR controller manages to provide this dynamics, the PI controller produces an undershoot and converges slowly to the target value.

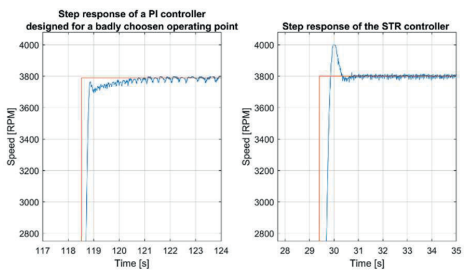


Fig. 21: Output signal settling STR controller vs PI controller.

The PI controller does not perform as well as in the middle of the operating range and cannot provide the slight overshoot which is required. The closed-loop system does not have enough gain and the integral action of the controller has to compensate the undershoot. However, the integrator is not fast enough, it takes up to four seconds to reach the target value, which is approximately two times slower than the STR controller. Although this effect is not very spectacular, it demonstrates the disadvantages of a static regulator designed for a single working point. If a system has mild non-linear characteristics, its identification is not accurate, or its properties are time dependent, a static controller will not be able to deliver the desired closed-loop response for all operating conditions. That can cause serious problems for high precision systems.

The results of the operating point variation tests suggest that the parameters of the target system are depending on operating point and time. Therefore, a second test is conducted to identify time dependence. Since the windup protection enables the estimation only after significant target value changes and only until the target value is reached it is not an option to simply set a target value and let the system run for a longer period of time. Periodically disabling the windup protection is also no option because this would lead to an estimation windup again [1]. Therefore, it is decided to choose an operating point of 1900 [RPM] and to periodically toggle the target value in a range of ±200 [RPM] for twenty minutes.

The identified parameters of this experiment clearly reveal that the system parameters are time dependent. Both numerator and denominator coefficients significantly change immediately after the start of the experiment and it takes them almost four minutes to converge to a final value range. Although the denominator parameters remain nearly constant after finishing converging, the numerator coefficients vary significantly around their final value. This implies that although the dynamics of the system remains almost constant, the system gain is permanently changing even for only slightly varying operating points (1900 ± 200 [RPM]). The change of coefficients during the startup phase is considered as a consequence of the system warm-up and heat transfer. Apparently it takes four minutes for the motor to reach the operating temperature. It is assumed that after reaching the steady state the temperature variations have only a minor influence on parameter changes. Knowing the properties of the laboratory setup [3] the friction is considered to play a bigger role in the parameter changes at the operating temperature. Generally, it can be declared that system gain is permanently changing even for a constant operating point. The experiment proves, that the parameters of the system are not only time dependent, but they also change with respect to time. Nonetheless, the STR controller managed to track those changes and to design a proper controller.

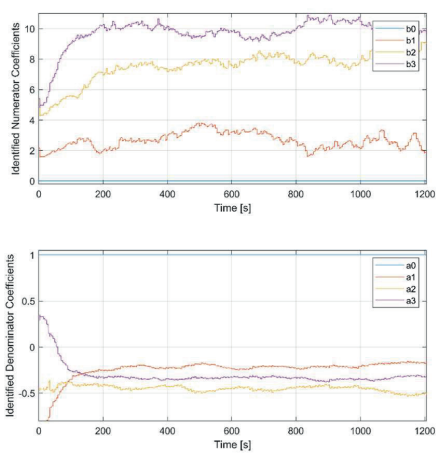


Fig. 22: Development of the identified system parameters during long time test.

It has to be noted that the selection of the observer polynomial plays a significant role in the case of the second test as well. Executing the same test with the initially chosen observer polynomials leads to a partly unstable result. Although the experiments optimized with the trial and error observer polynomial produced more stable results it can not be guaranteed, that this system remains stable for all operating points.

5. RESULTS

This thesis shows how an STR controller can be designed efficiently. Furthermore, it shows that a well designed STR controller can successfully identify the changing properties of the system in real-time, it is able to design a controller such a way, that the closed-loop system has the defined behaviour in all operating point and is able to dynamically switch its own controller parameters without getting unstable. With these results the original purpose of this thesis is fulfilled. However, during the development of the project several topics were identified, which would need further research. The sensibility against noise is identified as a significant drawback of the STR control. Therefore, it would be advantageous to elaborate a reliable and straightforward rule for the determination of the observer polynomial. The trial and error method delivers acceptable results in work. However, it is not proven mathematically that the system can never get unstable. With increasing model complexity the application of the trial and error method will become more difficult or even impossible. Even if it would be manageable to find the observer coefficients that way, it is not considered to be a reliable solution for industrial applications with high quality standards and strict security requirements. As an alternative, some kind of numerical solution combined with a proper noise characteristic simulation is suggested. Considering the laboratory setup used it turns out that the control can also be realized by a static PI controller in acceptable quality, hence the usage of an STR controller is not indispensable. However, the non-linear characteristics of the system would suggest the usage of an adaptive controller. Since the model structure has a moderate complexity the Gain Scheduling method could be tested as an alternative to the STR method. Generally, it would be interesting to test all adaptive control methods presented, to find out which one has the most benefits and the least design efforts. During the simulations it turned out that the STR control scheme can have difficulties with non-minimum phase systems. However, this problem is not further examined, because the laboratory setup could be stabilized by choosing a proper observer polynomial. Nonetheless, it would be interesting to find out under which circumstances the STR method would fail. In summary, it can be claimed that the thesis achieved its originally defined goals and could even highlight several additional research potentials, which exceeded the scope of this work.

6. CONCLUSION

Compared to a static PI controller, the STR controller provides more flexibility and operates with nearly the same efficiency. However, when used to control the examined laboratory setup except for some situations at the borders of the operating range, the PI controller does not show major drawbacks when compared to the STR controller. In these cases, the advantages of the STR controller are clearly visible as in contrast to the PI controller it can produce the desired system dynamics. In general, however, there is no significant difference between the efficiency of the two controllers, which means that although the laboratory setup has non-linear characteristics and its parameters are time-dependent, its control can be satisfactory realized with a PI controller.

7. REFERENCES

[1] K. J. Åström and B. Wittenmark, *Adaptive control*. Reading, Mass.: Addison-Wesley, 1995.

[2] J. F. Gieras, *Permanent magnet motor technology: Design and applications,* 3rd ed. Boca Raton: CRC Press, 2010.

[3] amira GmbH, “amira - DR300 Laboratory Setup Speed Control with Variable Load,” Technical Datasheet, Bismarckstraße, 67, D-47057 Duisburg, Germany, 2000. Accessed: Jul. 31 2022. [Online]. Available: https://docplayer.net/46966055-Dr300-laboratory-setup-speed-control-with-variable-load.html

[4] O. Nelles, *Nonlinear system identification: From classical approaches to neural networks, fuzzy models, and Gaussian processes*. Cham, Switzerland: Springer, 2021.

[5] G. Schulz and K. Graf, *Mehrgrößenregelung, Digitale Regelungstechnik, Fuzzy-Regelung Regelungstechnik*. München: Oldenbourg Wissenschaftsverlag, 2013.

[6] D. W. Clarke, “Self-tuning control of nonminimum-phase systems,” *Automatica*, vol. 20, no. 5, pp. 501–517, 1984, doi: 10.1016/0005-1098(84)90003-7.

ZOLTÁN BEK



Zoltán Bek BSc.

Fachhochschule Wiener Neustadt, Wiener Neustadt 2700, +4367761029902, zoltan.bek@fhwn.ac.at

I was born in 1998 in Sopron, Hungary. In the year 2016 I moved to Austria and made my Bachelor’s studies at the Mechatronics Faculty of the Fachhochschule Wiener Neustadt. I have been working at the company ZKW Elektronik GmbH since 2019 as an Embedded Software Engineer. There I develop real-time software in C and Simulink to the electrical control units of intelligent automotive headlamps.

BERECHNUNG DER SCHMIERFILMDICKE IN EINEM KEGELROLLENLAGER MITTELS EINES DATENBASIERTEN MODELLS

Gigov Boris

Kurzfassung: *Diese Arbeit schlägt ein datenbasiertes Modell vor, in welchem die Parameter der Hamrock-Dowson Gleichung mittels Ausgleichsrechnung anhand kleinster Fehlerquadrate an Messungen angepasst werden, um die Schmierfilmdicke in einem axialen Kegelrollenlager abzuschätzen. Das Systemverhalten wird mit einer dynamischen Analyse der klassischen Mechanik beschrieben. Zusätzlich wird dabei die Hamrock-Dowson Gleichung als Modellgleichung herangezogen, um die elastohydrodynamische Schmierung im Kontaktpunkt zwischen Rolle und äußerer Laufbahn zu beschreiben. Die Modellgleichung wird parametriert und die Parameter werden mittels kleinster Fehlerquadrate anhand bestehender Messungen identifiziert. Dieses Modell wird im Anschluss mit Hilfe der Gauß'schen Fehlerfortpflanzung validiert. Die vorgestellte Methode bildet die Basis zur Messung der Schmierfilmdicke in einem Kegelrollenlager anhand von Messungen der Last, Drehzahl und Temperatur. Die rasche Auswertbarkeit der Modellgleichungen des Modells ermöglicht in Zukunft die Echt-Zeit-Messung bzw. Prädiktion der Schmierfilmdicke in Kegelrollenlagern*

Schlüsselwörter: *EHL, Hamrock-Dowson, Schmierfilmdicke, Schätzung, Kegelrollenlager, kleinste Fehlerquadrate*

1. EINLEITUNG

Die vorliegende Arbeit ist als Teilprojekt einem Projektportfolio des AC2T zugeschrieben. Dessen übergreifendes Ziel ist es, den digitalen Zwilling eines Prüfstandes für Schmierstoffe zu entwickeln. Dabei sollen Kegelrollenlager in Echtzeit überwacht und der Degradationsgrad des im Lager verwendeten Schmierstoffes vorausgesagt werden. Der Alterungszustand des Schmierstoffes wird dann beurteilt, um einen frühzeitigen bzw. verspäteten Austausch vorzubeugen. Dadurch sollen Ressourcen gespart und Wartungsintervalle obsolet gemacht werden. Hierzu sind, neben anderen wissenschaftlichen Publikationen, auch seitens AC2T bereits Vorarbeiten geleistet worden. So steht bis dato ein moderner, an die Norm DIN 51350-6:1996-08 [1] angelehnter, Prüfstand für die Schmierstoffprüfung zur Verfügung. Zudem sind auch die Synchronisierung und Datenverarbeitung der verwendeten Sensoren von AC2T bereits durchgeführt worden.

Um dieses Vorhaben umzusetzen, werden folgende Methoden kombiniert bzw. angewandt:

- Die Hamrock-Dowson Gleichung [2]–[5] ist eine numerische Gleichung aus den 70er Jahren des 20ten Jahrhunderts, die eine Abschätzung für die Schmierstofffilmdicke in einem Punktkontakt zweier Körper ermöglicht. Ihre Formulierung mittels dimensionsloser Größen und die Tatsache, dass sie die Hertz'sche Pressung und Reynolds Gleichung für viskose Fluide vereint, machen die Hamrock-Dowson Gleichung zu einem fundamentalen Modell, um die Schmierung im elastohydrodynamischen Bereich (EHL) zu beschreiben.
- Das Newton-Euler-Schnittverfahren [6] wird angewandt, um über Geometrie des Kegelrollenlagers die kinematisch und kinetisch relevanten Größen zu berechnen. Es werden die Winkelgeschwindigkeit einer einzelnen Kegelrolle und die Kraft zwischen dieser und der äußeren Lagerlauffläche benötigt.
- Die Ubbelohde-Walther Gleichung [7] beschreibt die isobare Änderung der dynamischen Viskosität eines Schmierstoffes in Abhängigkeit der Temperatur bei atmosphärischem Druck.
- Mittels der Methode der kleinsten Fehlerquadrate [8] werden die vorab definierten Parameter der Modellgleichung anhand der Filmdickenmessungen identifiziert. Diese Parameter sind charakteristisch für die verwendeten Schmierstoffe und können in weiterer Folge für eine ex-situ Messung der Filmdicke der Schmierung verwendet werden.
- Die Gauß'sche Fehlerfortpflanzung [9], [10] wird genutzt, um den geschätzten Fehler des Modells zu analysieren und mit dem tatsächlichen Fehler zwischen gemessener und geschätzter Filmdicke zu vergleichen

2. PROBLEM- UND AUFGABENSTELLUNG

Die grundlegende Idee zur Modellierung der Schmierfilmdicke ist, eine Modellgleichung zu parametrieren anhand welcher man die Schmierfilmdicke innerhalb der Kontaktzone zweier Körper beschreiben kann. Die freien Parameter werden dabei mittels empirischer Messungen identifiziert werden. Der Informationsfluss für die Parameteridentifikation ist in Abb.1 veranschaulicht.

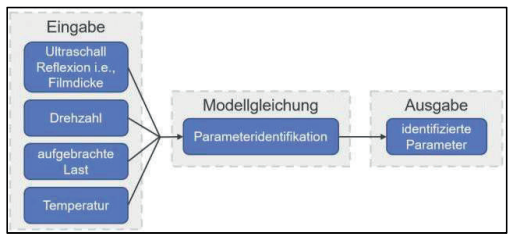


Abb.1: Informationsfluss bei der Parameteridentifikation

In weiterer Folge soll das parametrierte Modell die Schmierfilmdicke voraussagen können. Dadurch wird die direkte Messung der Filmdicke optional und man kann anhand der dynamischen Größen wie Drehzahl, Last und Temperatur Rückschlüsse auf den Schmierzustand ziehen, siehe Abb.2.

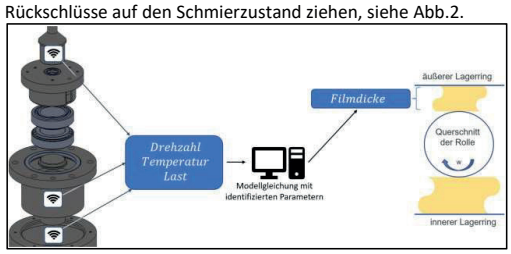


Abb.2: Prädiktion der Schmierfilmdicke mittels Modellgleichung

3. MATERIALIEN UND METHODEN

Da die beobachtbaren dynamischen Größen, Drehzahl, Last und Temperatur, lediglich das Gesamtsystem des Versuchsaufbaus beschreiben, werden zunächst die Normalkraft und die Tangentialgeschwindigkeit in einem ausgewählten Punkt des Rollenmantels bestimmt. Ähnliche mechanische Analysen wurden bereits unter Anwendung der Newton-Euler Schnittmethode von Gupta [14] durchgeführt. Im Falle der vorliegenden Arbeit wird aber eine vereinfachte Bewegung betrachtet, welche an eine Sonderform des Kollergangs erinnert. Hierzu wird angenommen, dass die Rolle und die Lagerringe starre Körper sind. Außerdem wird zwischen der Kegelrolle und der Lauffläche des Lagerringes keine Relativgeschwindigkeit zugelassen, d.h. es wird reines Rollen angenommen.

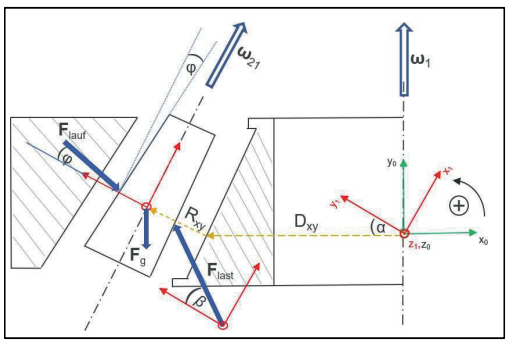


Abb.3: Mechanisches Modell eines SKF32008X Kegelrollenlagers im Querschnitt.

Abb.3 skizziert das vereinfachte mechanische Modell eines SKF32008X [15]. Dabei wird der Querschnitt des Lagers betrachtet. Das mitbewegte Bezugssystem $x_1 - y_1 - z_1$ ist mit dem Winkel α geneigt, wobei die Achsen des Bezugssystems in Richtung der Hauptträgheitsachsen der Kegelrolle zeigen. F_{last} stellt die Last auf ein einzelnes Rollelement dar und F_{lauf} die Normalkraft, welche zwischen äußerer Lagerlauffläche und dem Kegelrollenmantel wirkt, zudem sei F_g die Kraft der massebedingten Erdbeschleunigung der Kegelrolle. D_{xy} und R_{xy} bezeichnen den Durchmesser des inneren Lagerringes und den Radius der Kegelrolle. Die Nomenklatur wurde so gewählt, um das Lesen der nachstehenden Bewegungsgleichungen zu erleichtern, sodass man die Einflüsse der jeweiligen Geometrie eindeutig der Bewegung zuordnen kann. Der Winkel φ ist der Berührungswinkel der Kegelrolle und β der Angriffswinkel der Last F_{last}. Schließlich beschreiben ω_1 und ω_{21} die Winkelgeschwindigkeit des Bezugssystems und die relative Winkelgeschwindigkeit der Kegelrolle.

Da es sich hierbei um ein axial belastetes Kombinationslager handelt und sich die hydraulisch aufgetragene Last gleichmäßig auf die Anzahl der Rollen im Lager aufteilt, bezeichnet $F_{last} = F_{ges}/23$ den Anteil der Gesamtlast auf jeder 23 Rollen.

Anhand der somit festgelegten Geometrie kann die Kinematik des Systems berechnet werden. Hierbei wird ein Punkt an der Mantelfläche der Kegelrolle betrachtet, welcher in Kontakt mit der Lauffläche des äußeren Lagerringes steht. So wird die Absolutgeschwindigkeit $\vec{v}_{abs}$ des betrachteten Punktes im Koordinatensystem $x_1 - y_1 - z_1$, über die Summe der Führungsgeschwindigkeit $\vec{v}_F$ und der relativen Geschwindigkeit $\vec{v}_{rel}$ gefunden, sodass

$$v_{abs} = v_F + v_{rel} = \vec{\omega}_1 \times r_F + \vec{\omega}_{21} \times r_{rel} \tag{1}$$

gilt. Die Winkelgeschwindigkeit des Führungssystems $\vec{\omega}_1$ wird über die Drehzahl des Motors bestimmt und die relative Winkelgeschwindigkeit $\vec{\omega}_{21}$ sei zunächst unbekannt. Bildet man die Summe der Kreuzprodukte

$$\begin{aligned} v_F &= \vec{\omega}_1 \times r_F \\ &= \omega_1 \begin{pmatrix} \cos(\alpha) \\ \sin(\alpha) \\ 0 \end{pmatrix} \times \begin{pmatrix} -\frac{D_2}{2}\sin(\alpha) \\ 2r_2 + \frac{D_2}{2}\cos(\alpha) \\ 0 \end{pmatrix} \\ &= \omega_1 \begin{pmatrix} 0 \\ 0 \\ 2r_2\cos(\alpha) + \frac{D_2}{2}[\sin^2(\alpha) + \cos^2(\alpha)] \end{pmatrix} \end{aligned} \tag{2}$$

und

$$\begin{aligned} v_{rel} = \vec{\omega}_{21} \times r_{rel} &= \begin{pmatrix} \omega_{21} \\ 0 \\ 0 \end{pmatrix} \times \begin{pmatrix} 0 \\ (r_2) \\ 0 \end{pmatrix} \\ &= \begin{pmatrix} 0 \\ 0 \\ \omega_{21} r_2 \end{pmatrix}. \end{aligned} \tag{3}$$

Unter der Annahme, dass reines Rollen gilt, verschwindet die Geschwindigkeit v_{abs} im Berührpunkt des Kegelmantels mit dem Außenring,

$$\begin{aligned} v_{abs} = 0 &= v_F + v_{rel} \\ &= \omega_1 \begin{pmatrix} 0 \\ 2r_2\cos(\alpha) + \frac{D_2}{2} \\ 0 \end{pmatrix} \\ &+ \begin{pmatrix} 0 \\ 0 \\ \omega_{21} r_2 \end{pmatrix}, \end{aligned} \tag{4}$$

woraus ω_{21} bestimmt werden kann,

$$\omega_{21} = -\frac{2r_2\omega_1\cos(\alpha) + \frac{D_2}{2}\omega_1}{r_2}. \tag{5}$$

In weiterer Folge kann nun die Absolutbeschleunigung a_{abs} im betrachteten Punkt errechnet werden. Die zeitliche Ableitung der Absolutgeschwindigkeit v_{abs} (siehe Anhang) wird dabei zur besseren Übersicht wiederum als die Summe der Führungsbeschleunigung a_F, der Relativbeschleunigung a_{rel} und dem Coriolis-Anteil a_{cor} angeschrieben

$$a_{abs} = a_F + a_{rel} + a_{cor}. \tag{6}$$

Mit der absoluten Beschleunigung a_{abs} kann nun auf die Berechnung der Kinetik eingegangen werden. Hierzu werden die Zwangskraft F_{lauf} und die eingeprägten Kräfte F_{last}, F_g in den Schwerpunkt der Kegelrolle reduziert.

Schließlich stellt man das Kräftegleichgewicht im Schwerpunkt der Rolle auf

$$ma_s = \sum F, \tag{7}$$

Die somit gefundenen Winkelgeschwindigkeiten und Kräfte können als Einflussgrößen für die Hamrock-Dowson Gleichung (H-D) verwendet werden. Diese ist eine der wichtigsten numerischen Modellgleichungen zur Abschätzung der Schmierfilmdicke unter EHL-Bedingungen [16], [17]. Die Gleichung wird nun für die Zwecke dieser Arbeit parametriert, um die freien Parameter mittels kleinster Fehlerquadrate auf Basis der Messdaten zu wählen. Mit den so gefundenen Parametern soll die Modellierung eine alternative Bestimmung der Schmierfilmdicke in dem Kegelrollenlager ermöglichen, sodass eine in-situ Messung mit Ultraschall obsolet wird. Zunächst wird die H-D für die minimale Filmdicke wie folgt angeschrieben [18]

$$\frac{h_0}{R^*} = 3.63\left(\frac{u\eta_0}{E^*R^*}\right)^{0.68}\left(\frac{Q}{E^*R^{*2}}\right)^{-0.073}(\alpha E^*)^{0.49}(1 - e^{-0.68k}) \tag{8}$$

Wobei h_0 die minimale Schmierfilmdicke und u die mittlere Mitnahmegeschwindigkeit (eng.: „entrainment speed“) darstellt, mit welcher der Schmierstoff zwischen Kegelrolle und Lauffläche „eingesaugt“ wird. Q sei die Normalkraft und α ein von der Last abhängiger Piezo- bzw. Druck-Viskositätsfaktor, der die druckabhängige Viskosität des Schmierstoffes beschreibt. η_0 hingegen beschreibt die temperaturabhängige Viskosität bei einem Druck von $1bar$. Die H-D ist für den allgemeinen Fall einer Berührungsfläche zweier Kontaktkörper ausgelegt, die durch eine Ellipse beschrieben wird. Hierzu wird der Parameter für die Elliptizität k [3] eingeführt. Allerdings wird in dem hier präsentierten Modell der vereinfachte Fall einer Linienberührung herangezogen. In diesem Fall gilt $k \to \infty$ und mit $e^{-0.68k} \to 0$ reduziert sich der Klammerausdruck somit zur Identität. Schließlich beschreiben E^* und R^* das reduzierte E-Modul und den reduzierten Radius der sich berührenden Körper [19]. Da in Glg.(8) h_0 auf R^* bezogen wird, stellt sie in ihrer aktuellen Form, die in der Literatur oft verwendeten, dimensionslosen Größen dar [2]–[5], [16]–[18], wobei

$$H = \frac{h_0}{R^*} \tag{9}$$

der dimensionslose Schmierfilmdicken-,

$$U = \frac{u\eta_0}{E^*R^*} \tag{10}$$

der dimensionslose Geschwindigkeits-,

$$W = \frac{Q}{E^*R^{*2}} \tag{11}$$

der dimensionslose Last- und

$$G = \alpha E^* \tag{12}$$

der dimensionslose Materialparameter sind, sodass gilt

$$H = 3.63 \cdot U^{0.68} \cdot W^{-0.073} \cdot G^{0.49} \tag{13}$$

Unter Kenntnis von U, W und G kann nun die Parametrierung der Modellgleichung durchgeführt werden, wobei die Exponenten von Glg.(8) geschätzt werden,

$$H = 3.63^k U^a W^b G^c \ . \tag{14}$$

Dabei wird für den Faktor 3.63 ein eigener Exponent k definiert, um mittels Logarithmierens eine Linearisierung der Gleichung für die gesuchten Parameter k, a, b und c zu ermöglichen,

$$\log_{10} H = a \cdot \log_{10} U + b \cdot \log_{10} W + c \cdot \log_{10} G + k \cdot \log_{10} 3.63 \tag{15}$$

Die Ausführung der Identifikation mit Hilfe einer quadratischen Gütefunktion erfolgt in starker Anlehnung an O.Nelles [8]. Da für die Identifikation der definierten Parameter k, a, b und c Messdaten verwendet werden, wird Glg.(15) in vektorieller Schreibweise angegeben, wobei jede Zeile eine Messung bzw. Datenpunkt darstellt

$$\underbrace{\begin{pmatrix}\log_{10} H_0\\ \log_{10} H_1\\ \vdots\\ \log_{10} H_{n-1}\\ \log_{10} H_n\end{pmatrix}}_{y} = \underbrace{\begin{pmatrix}\log_{10} U_0 & \log_{10} W_0 & \log_{10} G_0 & \log_{10} 3.63\\ \log_{10} U_1 & \log_{10} W_1 & \log_{10} G_1 & \log_{10} 3.63\\ \vdots & \vdots & \vdots & \vdots\\ \log_{10} U_{n-1} & \log_{10} W_{n-1} & \log_{10} G_{n-1} & \log_{10} 3.63\\ \log_{10} U_n & \log_{10} W_n & \log_{10} G_n & \log_{10} 3.63\end{pmatrix}}_{X} \cdot \underbrace{\begin{pmatrix}a\\ b\\ c\\ k\end{pmatrix}}_{\vec{w}} . \tag{16}$$

Auf der linken Seite sind die bekannten gemessenen und logarithmierten Filmdicken, und auf der rechten Seite die bekannte logarithmierte Datenmatrix mit den verarbeiteten Sensordaten für Drehzahl, Last und Temperatur. Die Datenmatrix wird mit dem unbekannten Vektor der Modellparameter multipliziert. Um die unbekannten Parameter des unbekannten Gleichungssystems zu finden, muss mit einer Gütefunktion I eine Lösung für das Gleichungssystem gefunden werden,

$$I = \frac{1}{2}\sum_{i=1}^{n} e_i^2 = e^T e \cdot \frac{1}{2} . \tag{17}$$

I ist die Gütefunktion und e der Fehler des Modells. Dieser wird berechnet, indem man die Modellantwort $X\,\vec{w}$ (= der Schätzwert) vom realen Systemausgang y subtrahiert,

$$e = y - X\,\vec{w} \ . \tag{18}$$

Um nun $\vec{w}$ explizit darzustellen wird e in die Gütefunktion eingesetzt

$$\begin{aligned} I = \frac{1}{2}e^T e &= \left(\frac{1}{2}y - X\,\vec{w}\right)^T (y - X\,\vec{w}) \\ &= \frac{1}{2}\vec{w}^T X^T X\vec{w} - y^T X\,\vec{w} \\ &\quad + \frac{1}{2}y^T y \ , \end{aligned} \tag{19}$$

und das Minimum der Gütefunktion gesucht,

$$\frac{\partial I}{\partial\,\vec{w}} = X^T X\,\vec{w} - y^T X = 0 . \tag{20}$$

Aus dieser Bedingung kann $\vec{w}$ bestimmt werden,

$$\vec{w} = (X^T X)^{-1} X^T y \ , \tag{21}$$

sodass gilt

$$\begin{pmatrix} a \\ b \\ c \\ k \end{pmatrix} = (X^T X)^{-1} X^T y \,, \tag{22}$$

wobei der Ausdruck $(X^T X)^{-1} X^T$ als Pseudoinverse bezeichnet wird. Diese wird schließlich und den Filmdickenmessungen y multipliziert. Es sei an dieser Stelle betont, dass die Argumente 3.63 und G der Exponenten k und c hier als konstant angenommen werden. Somit wären die letzten beiden Spalten in der Datenmatrix X in Glg.(16) ebenfalls konstant, womit eine Lösung des Gleichungssystems nicht möglich ist, da die Datenmatrix keinen vollen Rang hätte. Daher werden 3.63 und G unter einem gemeinsamen Exponenten c^* zusammengefasst

$$H = U^a W^b (3.63G)^{c^*} \tag{23}$$

und mit dem Gleichungssystem

$$\begin{pmatrix} a \\ b \\ c^* \end{pmatrix} = (X^T X)^{-1} X^T y \,, \tag{24}$$

die Parameter a, b und c^* gefunden.

Allerdings muss die Genauigkeit des Modells ebenfalls abgeschätzt werden. In [8] wird dazu eine Methode gezeigt, bei welcher zunächst die Kovarianzmatrix der Parameter $cov\{w\} = p$ berechnet wird, um schließlich die Standardabweichung des Modells $std\{Xw\}$ zu schätzen. So wird zunächst das Rauschen geschätzt, indem der Fehler Glg.(18) quadriert, dann gemittelt

$$\hat{\sigma}^2 = \frac{e^T e}{N - n} \tag{25}$$

und schließlich mit der inversen Hessematrix $(X^T \cdot X)^{-1}$ multipliziert wird

$$p = \hat{\sigma}^2 \cdot (X^T \cdot X)^{-1}, \tag{26}$$

um die Kovarianzmatrix der Parameter zu erhalten $p = \begin{pmatrix} p_{aa} & p_{ab} & p_{ac^*} \\ p_{ba} & p_{bb} & p_{bc^*} \\ p_{c^*a} & p_{c^*b} & p_{c^*c^*} \end{pmatrix} = cov\{w\}$. In ihrer Hauptdiagonale zeigt sie die Varianzen der Parameter selbst und abseits dieser die Kovarianzen jeweils zweier Parameter, also die Korrelation der Parameter zueinander. Nun kann die Genauigkeit des Modells als eine Schätzung für die Standardabweichung dargestellt werden, indem die Quadratwurzel aus dem Produkt des Datensets mit der Kovarianzmatrix berechnet wird,

$$std\{\hat{y}\} = \pm\sqrt{X \cdot p \cdot X^T}. \tag{27}$$

Hier ist anzumerken, dass die Datenmatrix X und ihre transponierte X^T, jeweils ungefähr 290.000 Zeilen bzw. Spalten besitzen, da dies der Anzahl der Messungen pro Versuchsdurchgang entspricht. Hingegen ist die Kovarianzmatrix $cov\{w\}$ eine 3x3 Matrix. In allen Matrizen werden die Daten als Datentyp „float“, also mit 64Bit gespeichert. Wenn man nun Glg.(27) ausführen würde, bräuchte man einen RAM mit einem Speicherplatz größer als 673GByte, um mit allen Messpunkten rechnen zu können. Eine mögliche Alternative wäre, nur die Diagonalelemente der $cov\{w\}$ zur Rechnung heranzuziehen, da nur diese eine Aussage über die Varianz der Parameter tätigen. Somit wäre nur ein Drittel des Speicherplatzes, also ~224GByte, nötig. Mit der Gauß'schen Fehlerfortpflanzung [9], [10] ist allerdings eine Methode gegeben, anhand welcher man zuverlässig und effizient eine Schätzung für die Diskrepanz zwischen Modell und Messung finden kann. Sind die Eingangsgrößen nicht unabhängig, d.h. sie korrelieren, dann lautet die Schätzung σ_q^2 der Varianz für ein Modell der Form $q(x, y)$

$$\sigma_q^2 = \left(\frac{\partial q}{\partial x}\right)^2 \sigma_x^2 + \left(\frac{\partial q}{\partial y}\right)^2 \sigma_y^2 + 2\left(\frac{\partial q}{\partial x}\frac{\partial q}{\partial y}\right)\sigma_{xy} \,. \tag{28}$$

In Glg.(28) wird die Nomenklatur aus [10] verwendet, $\sigma_x^2, \sigma_y^2, \sigma_{xy}^2$ beschreiben die Varianzen der Variablen x, y bzw. deren Kovarianz. Da die Modellgleichung für die Schmierfilmdicke in Glg.(23) gegeben ist, und die Varianzen bzw. Kovarianzen wie in Glg.(26) berechnet werden, kann ohne weiteres die Fehlerfortpflanzung angeschrieben werden, sodass für die Varianz der modellierten Schmierfilmdicke $var\{h_{mod}\}$ gilt,

$$\begin{aligned} var\{h_{mod}\} &= \left[\left(\frac{a}{U} \cdot U^a W^b (3.63G)^{c^*} R^*\right)^2 \cdot p_{aa}\right. \\ &+ \left(\frac{b}{W} \cdot U^a W^b (3.63G)^{c^*} R^*\right)^2 \cdot p_{bb} \\ &+ \left(\frac{c^*}{G} \cdot U^a W^b (3.63G)^{c^*} R^*\right)^2 \cdot p_{c^* c^*}\left.\right] \\ &+ 2\left[\frac{ab}{UW}\left(U^a W^b (3.63G)^{c^*} R^*\right)^2 \cdot p_{ab}\right. \\ &+ \frac{ac^*}{UG}\left(U^a W^b (3.63G)^{c^*} R^*\right)^2 \cdot p_{ac^*} \\ &+ \left.\frac{bc^*}{WG}\left(U^a W^b (3.63G)^{c^*} R^*\right)^2 \cdot p_{bc^*}\right] . \end{aligned} \tag{29}$$

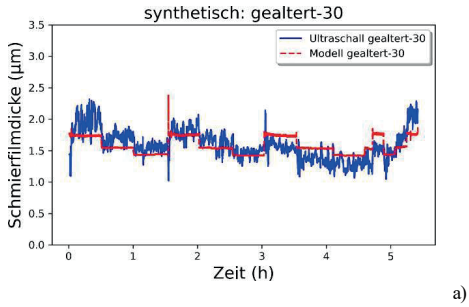
Die Modellgleichung kann als gemeinsamer Faktor herausgehoben werden und man erhält

$$\begin{aligned} var\{h_{mod}\} = R^{*2} \cdot U^{2a} W^{2b} (3.63G)^{2c^*} \\ \cdot \left[\left(\frac{a^2}{U^2} \cdot p_{aa} + \frac{b^2}{W^2} \cdot p_{bb}\right.\right. \\ + \frac{c^{*2}}{G^2} \cdot p_{c^* c^*}) \\ + 2\left(\frac{ab}{UW} \cdot p_{ab} + \frac{ac}{UG} \cdot p_{ac}\right. \\ \left.\left.\left. + \frac{bc}{WG} * p_{bc}\right)\right] . \end{aligned} \tag{30}$$

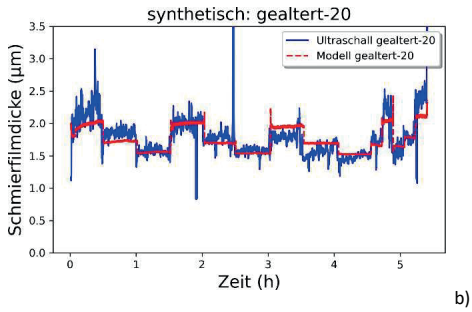
5. ERGEBNISSE

Die vorgestellten Methoden werden auf bereits bestehenden Messergebnissen angewendet. Diese wurden von Mitarbeitern des AC2T am Prüfstand abgenommen und aufbereitet. Geprüft werden dabei zwei Schmierstoffe, die hier als „synthetisch“ und „mineralisch“ bezeichnet werden, um sie durch das jeweilige Herstellverfahren zu unterscheiden. Außerdem werden die Schmierstoffe präpariert, um einen Alterungszustand zu simulieren. Die Alterungsstufen werden hier mit -30, -20, -10, frisch, +10, +20, +30 in 10er-Schritten dargestellt. Die Alterung wird durch eine Unter- bzw. Übersättigung des Schmierstoffes mit dessen eingelagerten Schmieröls simuliert. Dabei bezeichnen die Namen der Alterungsstufen den prozentualen Anteil einer Unter- bzw. Übersättigung bezogen auf den gesättigten Zustand, der als „frisch“ angeführt wird.

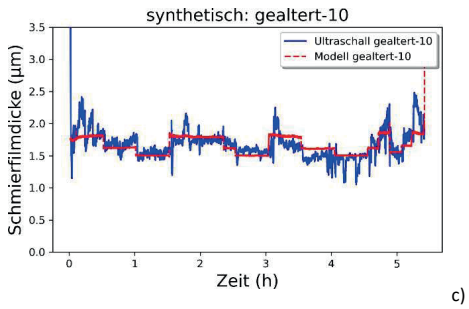
Jedes Präparat wird einem mehrstufigen Versuchsplan unterzogen. Dabei werden die Last, mit 5kN, 10kN, 15kN und die Drehzahl mit 1000U/min, 1500U/min, 2000U/min stufenweise variiert. Zeitgleich wird der Zustand des Versuchsaufbaus anhand von Messungen der Temperatur und des Drehmoments beobachtet. Die Versuche bei jeweils einer Alterungsstufe des Schmierstoffes dauern ungefähr 4,5 Stunden und werden in dieser Zeit mit ungefähr 290.000 Messpunkten dokumentiert.



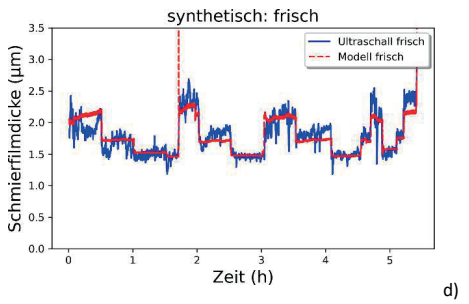
a)



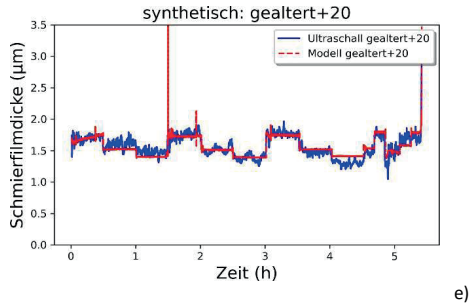
b)



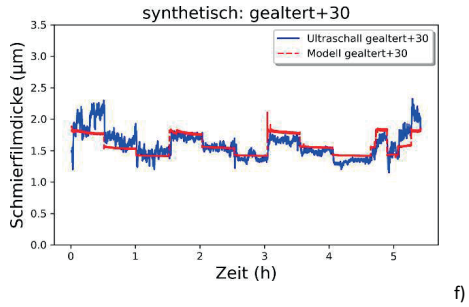
c)



d)



e)



f)

Abb.4: Parameterfit Messreihe „synthetisch“

Für die Messreihe „synthetisch“ sei erwähnt, dass die Ergebnisse für „synthetisch: gealtert+10“ nicht herangezogen werden, da die Messung vorab als nicht zuverlässig beurteilt wurde. Bei den restlichen Alterungsstufen Abb.4 zeigt sich, dass die H-D Glg. die Schmierfilmdicke sehr gut voraussagen kann. Dies bestätigt sich durch Berechnung des mittleren quadratischen Fehlers, nach Glg.(25). Der mittlere quadratische Fehler zeigt eine Abweichung in der Größenordnung von 0,1µm, was einem Fehler von <10%, bezogen auf die gemessene Schmierfilmdicke, entsprechen würde. Trotz der Übereinstimmung zwischen Messung und Modell, können einerseits Ausreißer seitens des Ultraschallsensors, dargestellt in Blau (Abb.4b), und andererseits Ausreißer seitens des Modells, dargestellt in Rot (Abb.4d und e), beobachtet werden. Es ist auch festzustellen, dass diese Abweichungen seitens des Sensors auch während einer Versuchsstufe auftreten können. Im Gegensatz dazu sind solche Diskrepanzen beim Modell nur an den Stufenübergängen zu beobachten. Der Grund dafür könnte im Versuchsaufbau selbst liegen, da derzeit keine Daten für die Systembeschleunigung vorliegen. Das heißt bei den Stufengrenzen des Versuchsvorgangs kommt es zu Übergangsbedingungen, die dem Modell nicht übergeben werden können, sodass eine Berücksichtigung der Winkelbeschleunigung nicht möglich ist. Es wäre möglich eine konstante Beschleunigung anzunehmen, allerdings würde dann der Beschleunigungsanteil der Bewegungsgleichung lediglich eine Konstante sein.

Die angeführten Ergebnisse lassen den Schluss zu, dass die Systemidentifikation mittels der Methode der kleinsten Fehlerquadrate erfolgreich ist und die gewählten Parameter der Modellgleichung bestimmt werden können, um eine verlässliche Prädiktion zu erzielen. Die mittleren Fehler zwischen den Messungen und den Prädiktionen belaufen sich auf <10% der Schmierfilmdicke. Dass Ausreißer wie in Abb.4b,d,e auf diesen Fehler kaum Einfluss nehmen liegt lediglich an der großen Anzahl der zur Verfügung stehenden Messungen. Wenn aber die Gauß’sche Fehlerfortpflanzung zur

Beurteilung der Modellgüte herangezogen wird, zeigt sich, dass alle Prädiktionen kritisch zu bewerten wären. Dem zu Grunde liegen die starken Abweichungen, welche bei den Variationen von Last und Drehzahl auftreten. Um diese Diskrepanz zwischen den berechneten mittleren quadratischen Fehlern und der Beurteilung durch die Fehlerfortpflanzung aufzulösen, müssten die Daten so gefiltert werden, dass die instationären Bereiche des Versuchsplans nicht einbezogen werden. Die Parametrierung der H-D Glg. scheint trotz allem ein guter Ansatz zu sein, da einerseits der mittlere quadratische Fehler zwischen Messung und Prädiktion relativ klein bleibt und andererseits sich das negative Vorzeichen für den lastbezogenen Parameter b mit der Literatur deckt [2]. Dort wird darauf hingewiesen, dass b stehts negativ sein muss. Dies wiederum würde bedeuten, dass das Modell je nach Parametertrend einen Gültigkeitsbereich aufweist, nämlich den Bereich in welchem $b < 0$ gilt.

6. ZUSAMMENFASSUNG UND AUSBLICK

Das Ziel dieser Arbeit ist eine Modellierung für die Schmierfilmdicke vorzuschlagen, die mittels Messungen eine Modellgleichung mit der Methode der kleinsten Fehlerquadrate so parametriert, dass die Schmierfilmdicke in einem Axial-Kegelrollenlager abgeschätzt werden kann. Der Vorteil dieses datenbasierten Modells ist es die Schmierfilmdicke in Kegelrollenlagern ohne direkte Messung mit einfach messbaren Eingangsgrößen wie Last, Drehzahl, Temperatur abzuschätzen, vorausgesetzt die Modellparameter werden auf den verwendeten Schmierstoff eingestellt. Damit ist eine aufwendige direkte Messung der Filmdicke nur im Labor nötig und die Anwendung im Feld wird immens erleichtert. Der experimentelle Zugang für den Laborversuch wird erläutert und die Methode der Ultraschallmessung der Schmierfilmdicke wird vorgestellt. In dieser Arbeit wird ein mechanisches Modell und die Parametrierung und Berechnung der Modellgleichung gezeigt. Die Ergebnisse zeigen, dass die Hamrock-Dowson Gleichung das Phänomen der elastohydrodynamischen Schmierung in einem Punkt sehr gut abbildet und das sogar mit als variabel angenommenen exponentiellen Parametern. Trotz des geringen Fehlers zwischen Messung und Modell (<10%), deutet eine Beurteilung der Standardabweichung der Ergebnisse durch die Gauß'sche Fehlerfortpflanzung auf einen Widerspruch hin, da sich die Varianzen nicht im gewünschten Bereich bewegen. Eine Erklärung dafür sind die Ausreißer bei den instationären Last- oder Drehzahlwechseln. Daher wird angenommen, dass das vorgestellte Modell nur im stationären Betriebszustand verlässlich ist, jedoch eröffnet es weiterführende Forschungsmöglichkeiten. So kann das mechanische Modell bzw. die Parametrierung der Modellgleichung ergänzt werden, um ein dynamisches Systemverhalten abzubilden, oder Echtzeit-Schätzungen mit Gewichtungs- bzw. Vergessensfaktoren zu implementieren. Weiters könnte der in python geschriebene Quellcode herangezogen werden, um eine Benutzerfreundliche Software zu programmieren. Es stünde auch in Aussicht eine Kooperation mit CAD-Software Herstellern ein Plugin zu designen, um Projekt-Assemblies in welchen geschmierte Maschinenelemente verwendet werden zusätzlich auf Lebensdauer der Lager zu simulieren und so virtuell zu prüfen.

7. LITERATURVERZEICHNIS

[1]Deutsches Institut für Normung, "DIN 51350-6:1996-08, Prüfung von Schmierstoffen - Prüfung im Shell-Vierkugel-Apparat - Teil 6: Bestimmung der Scherstabilität von polymerhaltigen Schmierölen." Beuth Verlag, Berlin, 1996. doi: https://dx.doi.org/10.31030/7197668.

[2]B. J. Hamrock and D. Dowson, "ISOTHERMAL ELASTOHYDRODYNAMIC LUBRICATION OF POINT CONTACTS - 1. THEORETICAL FORMULATION.," J Lubr Technol Trans ASME, vol. 98 Ser F, no. 2, pp. 223–229, 1976, doi: 10.1115/1.3452801.

[3]B. J. Hamrock and D. Dowson, "ISOTHERMAL ELASTOHYDRODYNAMIC LUBRICATION OF POINT CONTACTS - 2. ELLIPTICITY PARAMETER RESULTS.," J Lubr Technol Trans ASME, vol. 98 Ser F, no. 3, pp. 375–383, 1976, doi: 10.1115/1.3452861.

[4]B. J. Hamrock and D. Dowson, "Isothermal Elastohydrodynamic Lubrication of Point Contacts: Part III-Fully Flooded Result," J Tribol, vol. 99, no. 2, pp. 264–275, 1977, doi: 10.1115/1.3453074.

[5]B. J. Hamrock and D. Dowson, "Isothermal elastohydrodynamic lubrication of point contacts: Part IV—starvation results," J Tribol, vol. 99, no. 1, pp. 15–23, 1977, doi: 10.1115/1.3452973.

[6]P. K. Gupta, "Dynamics of Rolling-Element Bearings—Part I: Cylindrical Roller Bearing Analysis," Journal of Lubrication Technology, vol. 101, no. 3, pp. 293–302, Jul. 1979, doi: 10.1115/1.3453357.

[7]N. Gebhardt and J. Weber, Eds., Hydraulik – Fluid-Mechatronik. Berlin, Heidelberg: Springer Berlin Heidelberg, 2020. doi: 10.1007/978-3-662-60664-3.

[8]O. Nelles, Nonlinear System Identification. Cham: Springer International Publishing, 2020. doi: 10.1007/978-3-030-47439-3.

[9]U. Hahn, Elektrizität und Magnetismus, Optik, Messungen und ihre Auswertung. DE GRUYTER, 2015. doi: 10.1515/9783110377231.

[10]John R. Taylor, An Introduction to Error Analysis: The Study of Uncertainties in Physical Measurements, Second Edition. Sausalito: University Science Books, 1997.

Boris Gigov, BSc

FH Wiener Neustadt

BSc an FH Technikum Wien absolviert. Bachelorarbeit bei Semperit Technische Produkte GmbH in Wimpassing geschrieben. Ein Jahr Industriepraktikum bei Infineon in Villach, Kärnten. Masterarbeit bei AC2T in Wiener Neustadt verfasst.

HIGH-FREQUENCY IC DESIGN OF A COMMAND ENCODER FOR ENHANCED PARTICLE DETECTION IN MONOLITHIC ACTIVE PIXEL SENSORS AT ATLAS, CERN

Haberl S.

Abstract: *The ATLAS experiment is one of four major experiments at the LHC at CERN. Two particle beams are accelerated close to the speed of light before they collide. A collision of two particle beams is called an event. The ATLAS experiment consists of different layers of detectors to provide efficient tracking of the processes going on during an event. The detectors closest to the collision point are the pixel detectors. With the upcoming upgrade of the HL-LHC, the requirements increase a lot. New technologies must be considered to meet the requirements and reduce costs for the mentioned modules. This makes monolithic CMOS designs an interesting option. Those processes can be manufactured at high output rates and are much cheaper than non-standard processes. The latest approach is called MiniMALTA3 and is done in a 180 nm TowerJazz process.*

This thesis aims to design a digital module called Fast Command Encoder for this latest submission. This design's task is to provide a fast way of resetting and pulsing the pixels inside the matrix.

The design is done in RTL with synthesis and checked after PnR. The result is a Digital module working at 320 MHz and checked by simulation after PnR.

Keywords: *ATLAS, LHC, CERN, pixel detector, digital design, chip design.*

1. INTRODUCTION

Understanding the things around us is part of the human nature. During the past decades, a big step has been taken in describing matter that surrounds us by improving the standard model and fulfilling its predictions. One of the latest discoveries in that matter is the detection of the Higgs Boson and confirming the Brout-Englert-Mechanism. These observations were possible because the biggest particle accelerator ever built was used to research them, the Large Hadron Collider (LHC) at CERN. As particle generators transform energy into matter according to Einstein's equation E=m * c^2, bigger and bigger accelerators have to be used to generate more particles. The nature of some of those particles is, that they decay very fast, which makes them hard to observe. The probability of observing the decay of those particles can be increased by generating more particles in total, which means generating more energy in the accelerator. Another way of improving the detection of those particles is to increase the accuracy at which the decay products are measured. Some of those measurements are performed by silicon detectors, which can be found in all major experiments at CERN: ATLAS, ALICE, LHCb, and CMS [CERN 2021].

Silicon sensors, applied extensively on a daily basis, are well-received for their precision, sensitivity, and scalability. They have revolutionized the field of particle detection over the past decades. As time continues and technology improves, the detectors' requirements grow. The sensors improve and undergo changes on many levels to keep up with those. One desired goal is moving away from currently used sensors consisting of two silicon layers towards a monolithic design consisting of only one layer. Besides the advantages from a functional point of view, monolithic active pixel sensors (MAPS) show many benefits origin from the way they are built: less material, cheaper in production, and easier handling. This thesis aims to provide insight into the latest development of MAPS sensors at CERN [CERN 2021].

2. TASK

In order to provide a fast calibration of the MiniMALTA3, the inputs of the sensor module must change at high speeds. This requires for a digital module working at high frequencies such as 320 MHz. With an iterative calibration approach, reducing the time needed for a single iteration is essential. In addition to the calibration of the sensor the module is used to reset other modules in order to guarantee a seamless workflow.

3. METHODES

The design of the block is split up into several parts. The first part of the design is the description of the functionality in the register-transfer level (RTL). In this step, the desired logic (specified through the requirements) is written in *SystemVerilog*, and the functionality is verified by simulation. The testing is done in *CADENCE XCELIUM*. Based on the RTL the digital gates are picked during the synthesis, which match the functionality described in the previous step. The gates that can be used for this synthesis are defined by the technology used, the 180nm technology from TowerJazz. The selection of the different gates can be guided through constraints. Those constraints specify the desired current value at the output of the cells, the delay time on specific paths or the uncertainty margin of the arriving clock. This step is done via *CADENCE GENUS*. After the synthesis and with all the delays from different cells taken into account, the circuit is again simulated to verify the desired behaviour. The last step of the workflow is the actual placing of the digital cells on the chip. Based on this Place and Route process, the length, resistance, and capacity of the paths between the cells are added to the simulation. With these added parameters, the behaviour of the circuit on the chip can be approximated for a last simulation in *CADENCE XCELIUM*.

4. FAST COMMAND ENCODER

The FAST COMMAND ENCODER's (FCE) main purpose is to increase the calibrating speed of the sensor unit on the MiniMALTA3 (compared to the previous versions) as well as resetting its surrounding modules. It is designed in a 180 nm CMOS technology from TowerJazz.

4.1 Threshold Scan

To guarantee the desired behaviour regarding processing speed and noise of the read-out-electronics several parameters have to be adjusted. During a Threshold Scan (TS) electrons are inserted into the sensor via the circuit (this procedure is called pulsing) shown in Figure 1.

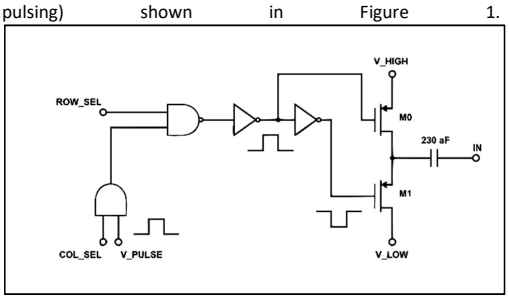


Figure 1: Circuit to insert electrons into the sensor

As the sensor is split up into multiple pixels consisting of several rows and columns, each pixel contains the circuit shown in Figure 1. The ROW_SEL and COL_SEL are handled by a different module called Slow Control (SC). Only V_PULSE is provided by the FCE. As the selection of the rows and columns can happen static, V_PULSE is the only dynamic component with influence regarding the execution speed. The setup of the sensor and the signals needed for the Threshold Scan is shown in Figure 2.

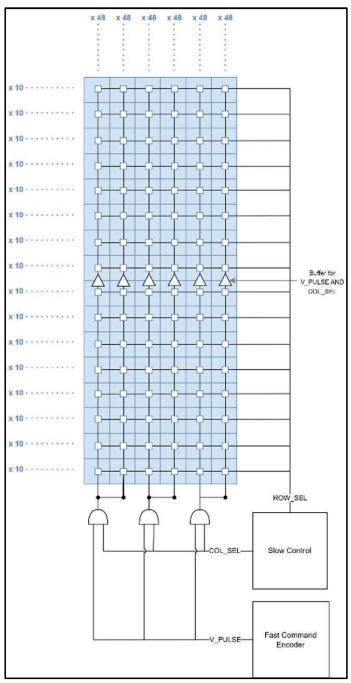


Figure 2: Connection of the FCE and the Slow Control to the sensor module

To display the need for faster Threshold Scans Figure 3 can be considered. It is easy to spot that the duration over an hour is not sufficient if multiple iterations are needed.

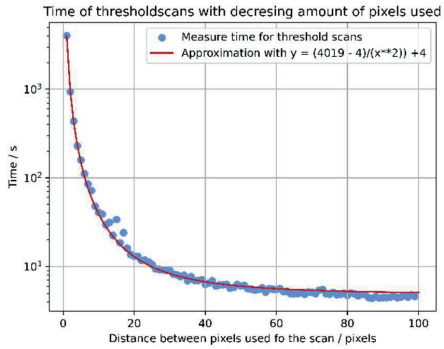


Figure 3: Correlation between the number of pixels and the duration of the threshold scan

4.1 Design

The FCE is controlled by an external FPGA and the SC. The FPGA is connected via a simplex connection, where predefined commands are send. As described previously, the two tasks of the FCE are resetting and pulsing, which also describe the FCE's outputs. Additional features are the variable duration of the output signals which can be selected by the SC and the deactivation of the FCE by the SC. A overview of several in- and outputs is given in Figure 4.

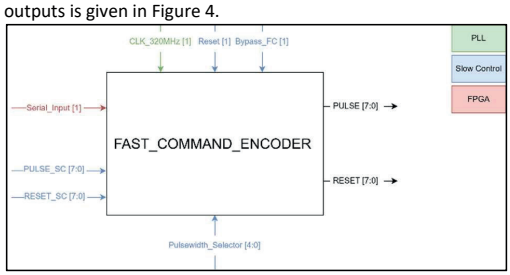


Figure 4: Block diagram of the FCE

The serial communication is started via a 8 bits starting sequence. Both commands (pulsing and resetting) require a different sequence:

PULSE: 8'b10101010
RESET: 8'b11001100.

A complete command consist of 24 bits, the 8 MSB display the starting sequence and the 8 LSB contain the actual data, the missing 8 bits are used as placeholder for further improvements.

4.2 Signal processing

The inputs from the FPGA is shifted into a 16 bit shift register. In the first stage, the starting sequence is analysed, to prepare the activation of the correct outputs. 16 clock cycles later the 8 LSB from the shift register are mapped to the output. The duration of 16 clock cycles is checked via counter, which is called *Counter Command*. The process of acquiring a command

is displayed in Figure 5.

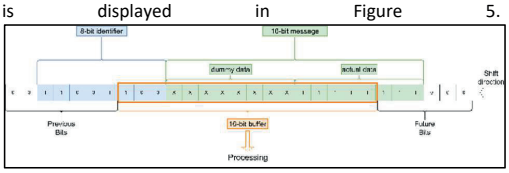


Figure 5: Processing of the incoming bit-stream from the FPGA

It takes one clock cycle to transfer the value from the shift-register to the output. At the same time as the transfer towards the output happens, the counter for the pulse width is enabled. After this single cycle, the shift register is reset at bit[15:2]. The last two bits are spared because in the moment of reset, a new bit has already arrived at the input, which would then be lost. This enables a continuous flow of commands toward the FCE. At the same time, when the shift register is reset, *Counter Command* is set back to 0. The pulse width given by the Slow Control is read every clock cycle until another counter (*Counter Pulse*) is activated. Once the outputs are activated, they are only reset after the defined amount of 320 MHz clock cycles. As *Counter Pulse* consists of 5 bits, the maximum duration of a single command would be 30 cycles. It takes 24 cycles to acquire a new command. A new command could arrive while an old one is still active. This behavior has to be considered during the operation. If no intersection between different commands is wanted, the maximum pulse width is 23 clock cycles, translating pulse lengths from 7.25 ns to 72 ns (93 ns with overlapping commands). This enables reset capabilities for modules down to 15 MHz synchronous reset. The pulse duration has no influence on modules with an asynchronous reset. At the minimum pulse width, the time between two pulses for a threshold scan is set at 75 ns.

4.3. Simulation and RTL

A typical case of an executed command is displayed in Figure 6. The shown command is guided towards four double columns. Columns 4 & 5, 6 & 7, 12 & 13 and 14 & 15 are pulsed in this example. The communication is started by the identifier for the command 8'b10101010 for pulsing, red line 1 in Figure 6. Once the whole identifier is received by the FCE, the identifier is stored in the *Command Identifier* register, red line 2. During this state, the next 16 bits are stored in the shift register, and the eight least significant bits are set as outputs with a one-cycle delay, line 3. The shift register and *Command Identifier* are reset one clock cycle after the final bit of data arrives, simultaneously when the outputs are set. The workflow is identical for pulsing and resetting with the difference that the reset signal is active. The chosen width must be applied at the input before processing the information is finished. This is possible because reading the input *Pulsewidth Selector* is faster than activating the flag that indicates an ongoing output. In either case, the desired wavelength should be applied before the transmission of the command is finished. As the pulse width is fixed when the output is set, *Pulsewidth Selector* can be changed during an active command. At the start of the command, the *Pulsewidth Selector* counts to the desired pulse width and resets the signal. If the width of the outgoing pulse is longer than 23 clock cycles, the output gets overwritten by the new command. In this case, Counter Pulse is not reset and continuous counting. This enables signals up to 31 clock cycles when the new command is identical to the previous one and permanently resets the output after 31 clock cycles.

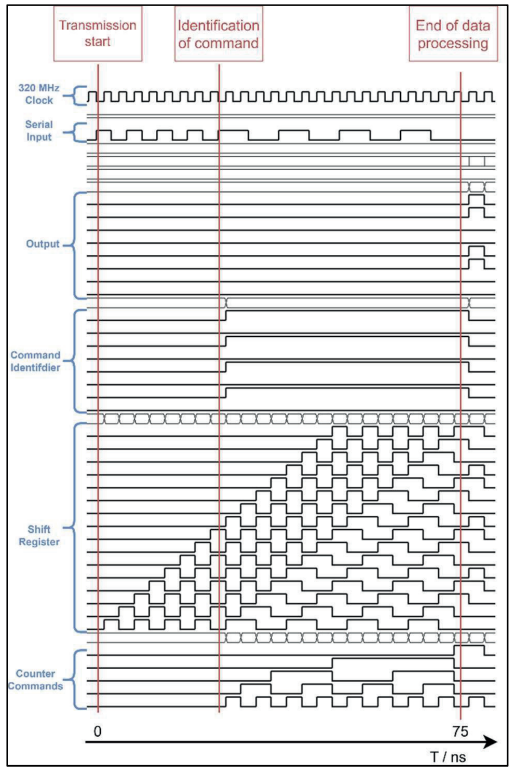


Figure 6: Typical workflow for executing a pulsing command

4.4. Synthesis

The first step of the synthesis is to define the constraints of the schematic according to the requirements. Starting from these constraints, *CADANCE GENUS* translates the functionality from code into digital components, according to the modules available in the used TowerJazz 180 nm technology. The final constraints applied are shown in Table 1.

Clock		
Max ramp time rising/falling edge	0.2	ns
Min ramp time rising/falling edge	0.2	ns
Clock setup uncertainty	0.55	ns
Clock hold uncertainty	0.55	ns
Output:		
Load	0.3	pF
False path	all	
Input:		
False path	all	
Max fanout	4	
Circuit:		
Max delay between *Input.RESET_SC* and *Output.RESET*	0.5	ns
Max delay between *Input.PULSE_SC* and *Output.PULSE*	0.5	ns

Table 1: Constrains applied for the synthesis of the FCE

The constraints regarding the clock describe imperfections that are to be expected from the clock signal during operation.

For example, 'Max ramp time rising/falling edge' describes the time needed to go from logic 0 to logic 1. This is required as wire-induced capacities delay the propagation of the signal during operation. These imperfections must be respected during the synthesis to make the circuit functional under real conditions. The setup time describes the time a signal has to be stable before the clock signal arrives in order to be processed correctly by a gate. For example, if the transition from logic 0 to logic 1 at the input of a flip-flop happens to be right before the clock arrives, the gate could treat the input equally as 0 or 1. On the other side, the hold time determines the time the input has to be stable after the clock arrives to guarantee defined behaviour [R. Munden 2005].

The load at the output essentially defines the driving strength of a module. It is determined as a capacity to overcome during a transition from 0 to 1 or inverse. This value, for example, corresponds to the length of the path it has to overcome or the number of cells it is connected to. The maximum fanout value describes the number of cells allowed to be connected to an output node. With strict timing constraints and the cells being at the limit of their driving strength, the number of connected cells is decisive. To connect more cells to the desired output node, the inputs of the following cells are wired to duplicates of the output cell. This operation comes at the cost of an increased usage of area on chip. The propagation time along this path can be set to a maximum value to define a critical delay between two nodes. This is especially important for time-critical signals and multi-bit signals. With signals consisting of more than one bit, it has to be guaranteed that all bits flip simultaneously. If this is not the case, a wrong bit combination could be provided for a specific amount of time. For example, this could result in unwanted behaviour in a state machine [Intel 2023].

5. RESULTS

During the Place and Route (PnR), the cells selected during the synthesis are placed in the actual layout of the MiniMALTA3 chip. With the final placement completed, the paths between different cells and the capacity between different wires can be calculated. Additional buffers are placed between different cells to overcome those distances and to guarantee optimal performance. Those buffers have an individual delay time each. Overcoming the actual signal paths also takes an additional amount of time. Those factors are considered in the simulation following the PnR. The delay between different pulsing signals is about 400 ps (see Figure 7b), and the timing difference between different resets is about 300 ps (see Figure 7a). These values must be considered during operation as some modules request a specific order of resetting.

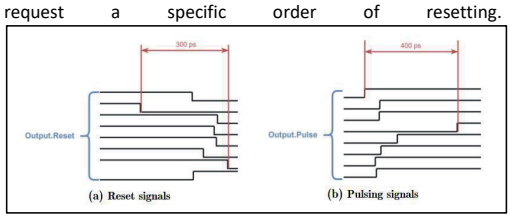


Figure 7: Delay between different signals from the same group after PnR

The final layout of the chip, after the PnR, is displayed in Figure 8.

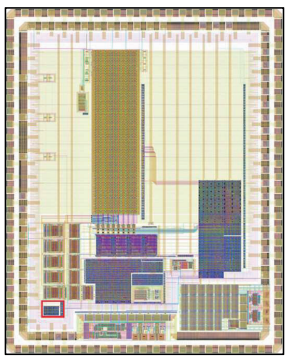


Figure 8: Final layout of the MiniMALTA3, the FCE is marked with a red rectangle

6. CONCLUSION

The results presented in this paper describe a digital module used for fast decoding of serial commands. With the fast decoding of commands, they can be executed in the order of ns. In this submission, the reset and pulsing commands can be executed with a 72 ns delay after the start of transmission. The design ran through extensive testing on the level of simulation.

7. LITERATURE

CERN. (2021, Jan.) Lhc the guide faq https://cds.cern.ch/record/2809109/files/CERN-Brochure-2021-004-Eng.pdf

Intel, *Intel® quartus® prime standard edition user guide* https://www.intel.com/content/www/us/en/docs/programmable/683492/18-1/constraining-designs.html

R. Munden, *ASIC and FPGA Verification*, ser. Systems on Silicon Ed. San Francisco: Morgan Kaufmann, 2005 https://www.sciencedirect.com/science/article/pii/B9780125115000



Sebastian Haberl

CERN, EP-ADE-ITK Sebastian.haberl@cern.ch

HTL-Matura 2017, Bachelor-Studium „Mechatronik und Mikrosystemtechnik“ an der FH Wiener Neustadt 2018 bis 2021, Master-Studium „Mechatronik“ an der FH Wiener Neustadt seit 2021

DEVELOPMENT OF A CLOSED LOOP HALL SENSOR FOR HIGH CURRENT MEASUREMENTS IN MVDC GRIDS

Michael Holzbauer

Abstract: *In this paper, the development of a closed loop Hall sensor for the current measurement range of 1kA-8kA in MVDC grids is described. Simulations of all necessary influencing parameters are carried out and analyzed. This is followed by the conceptual design for a prototype as well as temperature and accuracy measurements. Furthermore, additional amplifier electronics are developed in order to guarantee the required current compensation. Finally, a circuit board is produced, which can then be built into a new prototype. The respective tests regarding accuracy and temperature behavior are repeated, evaluated and then compared with the required project requirements.*

Keywords: MVDC, Hall Effect, Current measurement, Simulation, Closed-loop*.*

1. INTRODUCTION:

Current as well as voltage sensors have already been deployed and used in AC grids for decades. In contrast to AC grids, DC technologies are becoming more and more relevant. One of the reasons is that renewable energy sources tend to rely on DC technologies rather than AC. Therefore, DC technologies are emerging on the power supply side, for example via photovoltaic and on the consumer side (charging processes, DC motors, etc.). The advantages can also be derived from the example of a wind power plant. Wind power plants are usually installed in rural areas, which means that transmission to urban areas must be guaranteed. In this case, the power transmission is often done with HVDC. These offer the advantage of lower losses and higher efficiency compared to AC (Berthou, 2020). The progressive development of DC grids also increases the need for current and voltage sensors as well as protection mechanisms. The currently available sensors in the DC range, however, are only suitable for low-voltage applications. Within the framework of the EU-project Hyperride, a current sensor for medium voltage direct current (MVDC) grid is to be developed. The current sensor to be developed should offer temperature as well as offset compensation, immunity to external magnetic fields, high dynamic behavior as well as high accuracy over the whole measurement range which is between 1kA and 8kA. Furthermore, the sensor must withstand voltages up to 14kV. In order to develop the current sensor, different current measurement methods can be considered. When analyzing the different methods as well as their pros- and cons, the closed-loop principle was considered the most suitable for the development of the current sensor. Simulations and experimental setups were thus carried out. The gained knowledge was then used for the assembly and construction of the first prototype.

2. SCOPE:

The scope of this work deals with the implementation of a closed-loop Hall effect current sensor, which should withstand voltages up to 14kV, offer a high measurement range as well as good measurement accuracy across the measurement range. The main challenges are to meet the requirements regarding the desired accuracy and the isolation level.

3. MATERIALS AND METHODS:

In order to build a prototype as well as designing a signal conditioning circuit for the closed-loop principle, it is necessary to understand the theoretical background of the Hall effect as well as the closed-loop principle. The Hall sensor itself works on the basis of the Hall effect, which was discovered by Edwin Hall in 1879. The Hall effect occurs when a current carrying conductor is placed perpendicular to a magnetic field. By applying a voltage to the Hall element, current flows. Depending on the doping, either holes or electrons are used as charge carriers. The electron moves in the opposite direction of the current with an average velocity *v.* Due to the magnetic field, the electrons experience a magnetic force also known as Lorentz force. The electrons are thus deflected perpendicular to its direction of movement. This results in an electron surplus on one side and a shortage of electrons on the opposite side. Due to this charge separation, an electric field is generated, which counteracts the Lorentz force. The resulting voltage is then referred to as Hall Voltage. The Hall sensor thus generates an output voltage that is directly proportional to the strength of the magnetic field. The working principle is illustrated in figure 1.

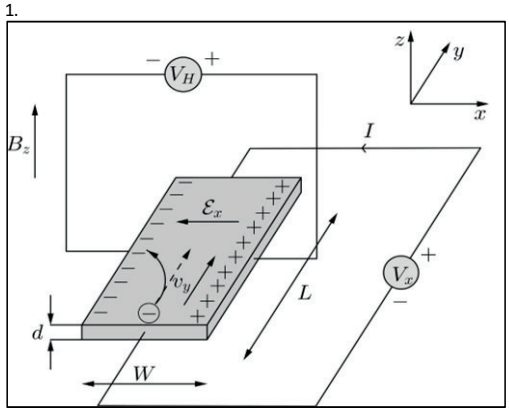


Figure 1: Hall effect (schematic) (Xanthakis, 2020)

The closed loop principle consists of an iron core, an airgap for the Hall-element, a current carrying conductor, a compensation coil, which is wrapped around the iron core and a signal conditioning circuit. The working principle is illustrated in figure 2. The compensation coil serves as negative feedback in order to compensate the magnetic flux in the airgap created by the current carrying conductor. The feedback path consists of the signal conditioning electronic and the compensation coil, which senses the magnetic field of the core and adjusts the current through the compensation coil with the aim of reducing the magnetic field inside the airgap to zero. The voltage drop across the shunt resistor can then be measured to determine the current being fed into the compensation coil.

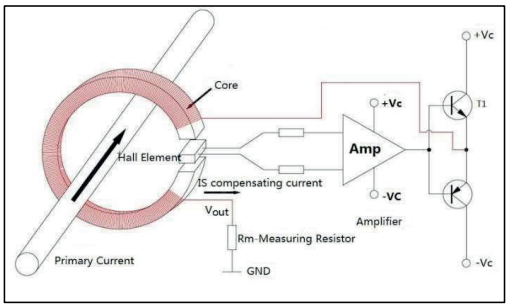


Figure 2: Basic closed-loop principle (Arar, 2022)

Equation 1 must be fulfilled to generate a magnetic flux density in the airgap of approximately zero.

$$B \ll \mu_0\mu_r * \frac{I_p * N_p}{l_a} \approx \mu_0\mu_r * \frac{I_c * N_c}{l_a}, \qquad (1)$$

Where I_p defines the primary current, N_p is the number of primary turns, N_p the number of secondary turns, I_c the compensation current and l_a defines the height of the airgap. As a first approach, FEM simulations were performed. The simulations were carried out with COMSOL 5.6/6.0. Simulations are particularly suitable for determining in advance whether the theoretical considerations are correct and comprehensible. Therefore, the Hall effect as well as the closed-loop principle is simulated. For the closed-loop principle, a 3D Model was set up, consisting of primary conductor, an iron core, a compensation coil, which was wrapped around the iron core and an airgap for of 2mm for the Hall element. Assuming that a primary current of 1000A is induced into the primary conductor and that the coil consists of 5000 windings a compensation current of 200mA would be necessary to reduce the magnetic flux density in the airgap to zero. The simulated result is shown below.

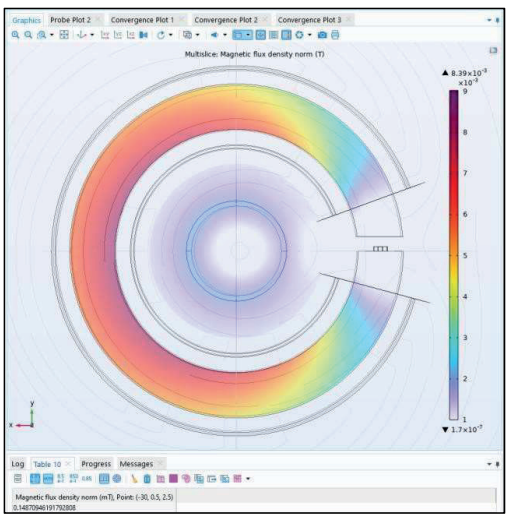


Figure 3: Simulated closed-loop approach

By evaluating the central point of the air gap, a magnetic field value of zero should be the case. In reality, air accumulates sporadically between the core and the compensation coil. Therefore, a leakage flux occurs, which leads to a distorted result. Since the used Hall element consists of Indium-antimonid (InSb), the material properties regarding the carrier concentration, hole- and electron mobility must be evaluated in order to specify the magnetic conductivity tensor. The charge conversation is then solved within the equations implemented in COMSOL. The current flow can be related to the electrical field via the anisotropic conductivity tensor, which is defined by a 3x3 matrix.

$$\begin{matrix} J_x \\ [J_y] = \\ J_z \end{matrix} \begin{bmatrix} \sigma_{xx} & \sigma_{xy} & \sigma_{xz} \\ \sigma_{yx} & \sigma_{yy} & \sigma_{yz} \\ \sigma_{zx} & \sigma_{zy} & \sigma_{zz} \end{bmatrix} \begin{matrix} E_x \\ [E_y] \\ E_z \end{matrix} \qquad (2)$$

Where J describes the current density, E symbolizes the electric field and σ is the material dependent conductivity. The anisotropic behavior is due to the Hall effect. The applied magnetic field B of the primary conductor produces the Lorentz force on the moving electrons. Depending on the magnetic flux density direction, different matrix terms for the material dependent conductivity result. The simulated output voltage as a function of the magnetic flux density is illustrated in figure 4.

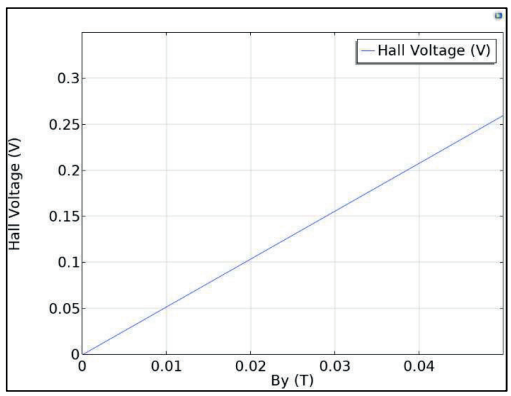


Figure 4: Simulated Hall voltage at Vc = 1V

To verify, whether the simulated Hall voltage corresponds to the actual indium antimonide Hall voltage, the data sheet of the InSb semiconductor Hall element was used and is visualized in figure 5.

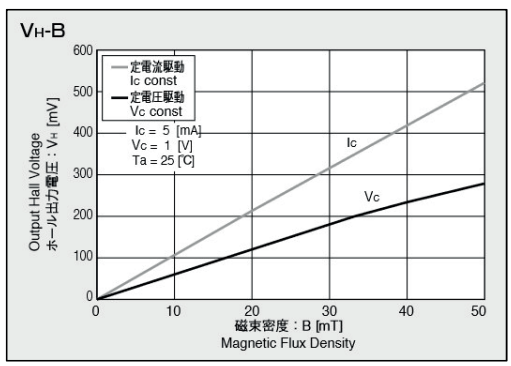


Figure 5: Hall voltage according to the datasheet (KASAI, 2023)

By comparing the two plots, visual similarities can be observed. The accuracy of figure 4 can be increased by choosing a finer mesh.

4. PRACTICAL IMPLEMENTATION:

In this section, different experimental setups are explained in detail in order to develop a first prototype as well as a PCB design. Since the Hall effect and the closed loop principle have already been explained in chapter 3, a first experimental assembly has been set up and tested. The setup consists of a conductor loop, an iron core, a compensation coil with 200 windings, the InSb Hall element and an evaluation module manufactured by Texas Instruments, which served as signal conditioning circuit. Since the maximum power of the evaluation module was limited to 250mW, the maximum current which could be measured was approximately 40A. Although, the project requirements cannot be fulfilled by using the evaluation module, an iron core was designed to measure 200A accurately. The following mathematical relationship had to be guaranteed in order to measure 200A accurately.

$$P_{board} \geq P_{shunt} + P_{idle} + P_{coil} \quad (3)$$

Where P_{boad} is the maximum performance of the board, P_{shunt} is the power consumed by the shunt resistor, P_{coil} is the power consumed by the compensation coil and P_{idle} is the power consumed by the board without driving current to the compensation coil. Apart from accuracy measurements, dielectric tests had to be performed to ensure the required isolation level of 14kV. The measurement setup of the dielectric test is illustrated in figure 6.

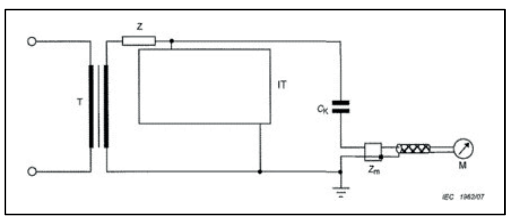


Figure 6: Partial discharge measurement setup (DKE, 1983)

For the measurement an EMC filter was integrated before the test transformer. Z symbolizes a filter, IT stands for the instrument transformer that has to be tested, C_k is the coupling capacitor, Z_m is the impedance of the measurement setup and M is the partial discharge measuring instrument. The actual parameter measured is the apparent charge in pico Coulomb [pC], which describes the quality of insulation. The upper and lower test points as well as the maximum voltage, that the prototype must withstand for one minute without taking any damage are declared in the standard IEC61869-2. During different accuracy measurements as well as during the construction process of the prototype a Hall offset was observed. According to (R. Portas, 2022), the most common reasons yielding an offset are:

- Magnetic influences
- Temperature influence
- Mechanical force (pressure)

Magnetic influences can be avoided by demagnetizing the iron core before an accuracy measurement is started. In order to estimate the influence on the offset due to changes in the ambient temperature, a climatic chamber test has been performed. The test result is analyzed in the section 5. Since the prototype is casted into epoxy resin, mechanical forces occur during the construction process. mechanical force yielding an offset. The setup is illustrated in figure 7.

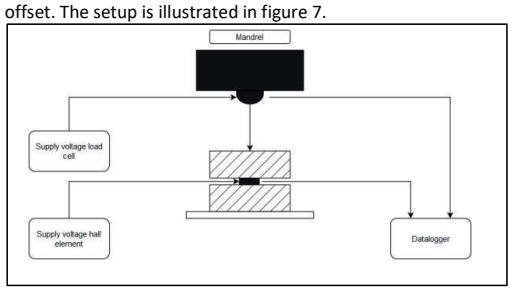


Figure 7: Mechanical force measurement setup (schematic)

The load cell was mechanically fixed in the mandrel. The shaded rectangles define wooden blocks that were positioned around the Hall element. The wooden blocks were used for various reasons. On the one hand, the Hall plate and placeholder are positioned in the air gap of the core. The mechanical forces generated during the curing of the epoxy resin would thus primarily affect the iron core and not the Hall element directly. Therefore, the wooden blocks represent the iron core surrounding the Hall element. Furthermore, without the wooden blocks, the Hall sensor would be destroyed within a very short time or the individual Hall element legs, which are necessary both for the supply and for the output signal, would break. The black rectangle between the wooden blocks thus defines the Hall element. In total, two different supply voltages are needed, one for the Hall element and one for the load cell. The measurement outputs of the load cell and the Hall sensor are then connected to the data logger. By rotating the spindle, the device together with the load cell moves in the direction of the Hall element. The force was applied centrally on the wooden block, which should minimize any kind of displacement primarily affecting the wooden blocks and the Hall element. To avoid possible displacement due to a non-optimally aligned load cell, wooden plates of approximately the same thickness as the placeholder of the Hall element were attached to the edge between the wooden blocks. The measured offset due to mechanical force is analyzed in section 6. The knowledge gained from the construction process as well as every aspect yielding an offset is used for the design of a new prototype. Since the project requirements demand, that the current sensor ensures a measurement range of 1kA up to 8kA an electronic circuit must be designed. Texas Instruments offers a test design, which is used to measure currents up to 100A accurately with the same integrated circuit, which is used in the evaluation module. The signal conditioning circuit is shown in figure 8.

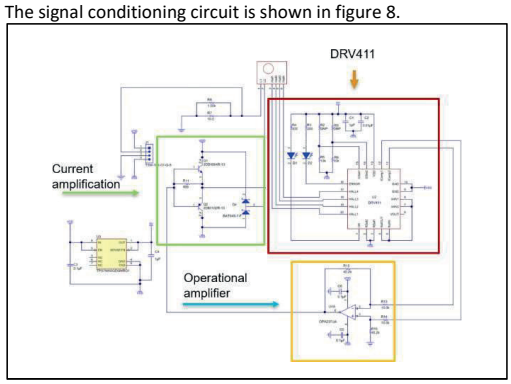


Figure 8: Electric circuit (schematic) (TI, 2020)

The circuit is divided into three different stages. The integrated circuit (DRV411) is used for temperature as well as offset compensation and resembles the first stage. Furthermore, it processes the output voltage of the hall sensor and drives an output voltage signal to the operational amplifier. The operational amplifier is the second stage and ensures that the used transistors operate in a favorable working range. The current amplification stage drives the current through one end of the compensation coil, generating a magnetic flux density, which serves as negative feedback and should therefore correspond to the magnetic flux density produced by the primary conductor. The second ending of the compensation coil is connected to a shunt resistor, in order to measure the current which is driven through the coil. The design visualized in figure 8, had to be adjusted to meet the defined requirements. More precisely, the operational amplifier gain has been adjusted and the transistors were exchanged to drive higher currents through the compensation winding. The results of the adapted design are visualized in section 6. Most of the accuracy measurements were performed with a breadboard/perfboard until a PCB-design was created. The final version of the PCB is illustrated in figure 9.

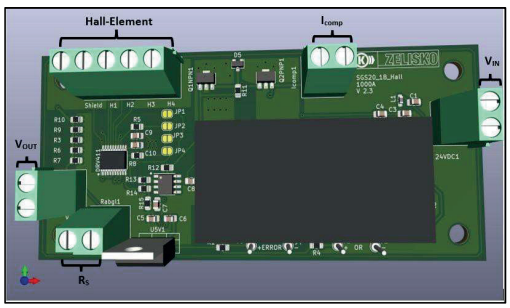


Figure 9: Final PCB-design

The Hall element and coil terminals are connected to the PCB through screw terminals. The other screw terminals are used for the output voltage, the shunt resistance and the supply voltage. The black rectangle on the PCB represents a DC/DC converter. The next section deals with the analysis of the different measurement setups.

5. RESULTS:

In this section, different measurement results regarding accuracy, isolation quality and offset investigation are analyzed. Figure 10 illustrates the first accuracy measurement using the evaluation module from Texas Instruments.

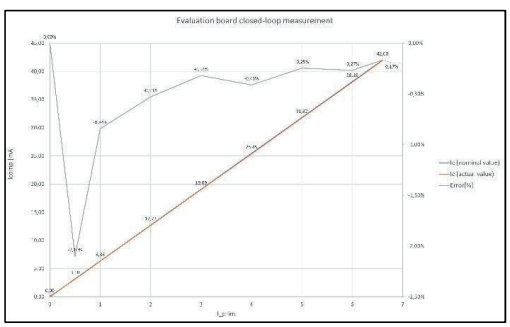


Figure 10: First accuracy measurement using the evaluation module

The first measurement with the evaluation board only measured the current through the coil I_c. The reason for this is that the functionality of the closed loop principle can be easily determined. For the curves shown in Figure 10, the nominal value for I_c was calculated and subsequently measured. The error between the measured value and the actual value is also shown in figure 10 as well. The nominal value was calculated by using the following equation:

$$I_{comp} = \frac{I_{prim} * N_p}{N_s} \quad (4)$$

Analyzing the error shows that the accuracy is increasing with higher primary currents. The evaluation board requires a supply voltage of 5V as well as a current limitation of 50mA. If the primary current is increased step by step, the permissible current in the evaluation board also increases. As soon as the 50mA is exceeded, an error conditions occurs. The supply voltage of the evaluation board collapses, while simultaneously trying to maintain the current amplification to the compensation coil in order to compensate the magnetic flux generated by the primary conductor. Nevertheless, at some point the board reaches its limit, where the magnetic flux, generated by the primary current can no longer be compensated and thus an output voltage is set. The second measurement result is visualized in table 1 and refers to the dielectric measurement setup visualized in figure 6.

Applied Voltage [kV]	PD-Value [pC]	Maximum permitted pC-value [pC]
12.2	2	20
17.5	2	/
21	2	50
38.5	2	/

Table 1: Partial discharge measurement results

Table 1 shows the measurement results of the built prototype. When interpreting the results, it is first necessary to note that the partial discharge test was performed with an AC voltage source. The reason for this is that DC partial discharge tests are not available. It is therefore assumed that the results obtained with AC are sufficient. Referring to the results obtained in Table 1, it is clear that the partial discharge value at each voltage value is 2pC and thus the test can be considered successful. The first and the third applied voltage value define the upper and lower test point. Since the dielectric measurement was successful, the results of the measurement setups with respect to possible Hall offsets are presented below. Figure 11 illustrates the results of the mechanical force measurement.

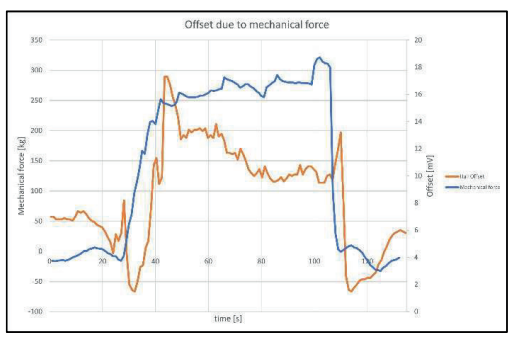


Figure 11: Generated offset due to mechanical force

When analyzing the measured values presented in figure 11, it became clear that mechanically generated pressure affects the offset of the Hall element. Furthermore, the applied force did not exceed 350kg. Since the Hall element is located in a 3D printed case, increasing the mechanical force might result in destroying the Hall element respectively the 3D printed case. In order to reduce the offset generated by mechanical force, the manufacturing process of the prototype has to be adjusted. As already mentioned in section 4, temperature affects the Hall element and produces an offset. The results of the climatic chamber test are thus illustrated in figure 12.

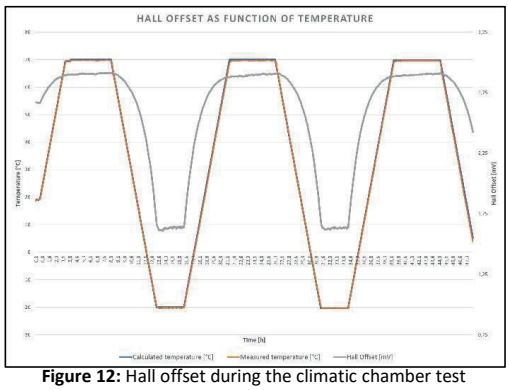


Figure 12: Hall offset during the climatic chamber test

The calculated temperature corresponds to the temperature measured with the temperature sensor. Furthermore, figure 12 shows that the Hall offset increases at high temperatures and decreases at low temperatures. Additionally, it is noticeable that the offset changes slightly at the constant temperature of +70°C and -20°C. In order to verify, that the results obtained in figure 35 are reproducible, another climatic chamber test was performed. This time the prototype was used instead of the uncasted Hall element. The temperature range as well as the measurement duration remained unchanged. In addition to the Hall offset, the coil resistance was measured. The analysis of the measured values revealed a similar behavior, but the offset exhibited an increase of 5mV over the entire temperature range. The higher offset is the result of the simultaneous resistance measurement of the compensation coil. During the resistance measurement, a small current is induced in the coil, which generates a weak field in the core, which in turn affects the Hall offset. This shows that in addition to temperature and mechanical influences, weak DC currents also cause magnetization of the core and a higher offset of the Hall element. Although the generated offset due to DC currents can be rectified by demagnetizing the iron core, the offset obtained from temperature influences remains. Since both measurement setups producing an offset have been investigated, the PDB design is analyzed. The final accuracy measurement result is hence illustrated in figure 13.

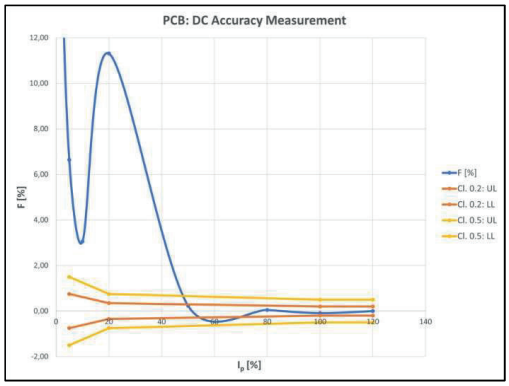


Figure 13: DC Accuracy result performed with the PCB

The measured values are furthermore displayed in table 2.

Ip [A]	Ip [%]	Vc [V]	Vm[V]	F [%]	Fc[%]
10	1	0.02	0.0248	36.40	36.66
20	2	0.04	0.0432	18.80	17.29
50	5	0.09	0.0982	8.02	6.64
100	10	0.182	0.190	4.39	3.06
200	20	0.364	0.410	12.75	11.31
500	50	0.909	0.923	1.53	0.24
800	80	1.455	1.47	1.34	0.05
1000	100	1.818	1.84	1.20	-0.09
1200	120	2.182	2.21	1.29	0.00

Table 2: DC accuracy values measured with the PCB

I_p defines the primary current of the DC current supply multiplied by the number of turns. F_c symbolizes the corrected error, which can be achieved by adjusting the shunt resistor. By adjusting the shunt resistor, the measurement error at the 100% point, or in this case at the 120% point, is reduced to 0. V_c is the calculated output voltage and V_m is the measured output voltage. Based on the results presented in table 2 as well as in figure 13, different optimization possibilities result, which are explained in detail in section 6.

6: SUMMARY AND OUTLOOK:

The present work deals with the realization of a Hall effect based current sensor designed for an isolation level of 14kV. Relevant factors such the closed-loop principle and the Hall effect were explained in more detail using FEM simulations in section 3. Based on the simulations, the practical implementation and construction of the first prototype was started. The first measurement setup as well as the development of the electronics were described. The accuracy results as well as the different factors yielding an offset have been analyzed. When comparing the project requirements with the measurements done in section 4 and 5, it becomes clear that the isolation level of 14kV is fulfilled. The accuracy of the prototype has been tested with AC and DC. The PCB design offers good accuracy for high currents but needs to be adjusted for lower currents in order to meet the desired accuracy class. Based on the measurement results obtained, various design possibilities can be considered. For example, the currently installed transistors can be replaced by more powerful ones. This would result in fewer turns being required, thus decreasing the coil resistance. Another possibility would be to increase the input voltage of the transistors to bridge the high resistance of the winding. Another approach would be to replace the bipolar transistors and use mosfets instead, which would likely lead to

different results in terms of accuracy, allowing for better results in the lower current measurement range as well. The following equations resemble different possibilities in order to optimize the current sensor.

$$I_{p,max} = \frac{V_{ss}}{R} \quad (5)$$

where V_{ss} defines the supply voltage for the transistors and R defines the combined resistance of the winding as well as the used shunt. $I_{p,max}$ symbolizes the maximum primary current which is led through the current sensor. By either increasing the supply voltage or decreasing the overall resistance, higher primary currents can be measured. The second correlation is given by equation 6.

$$I_s = \frac{I_p}{N'}$$

where Is is the compensation current, Ip the primary current and N the number of windings. In summary the requirements regarding the isolation level have been achieved. The desired accuracy is achieved for the nominal current of 1000A by adjusting the shunt resistance. Nevertheless, the measurement error across the complete measurement range needs to be improved. The Hall element offset can be lowered by adapting the construction/assembly process to reduce mechanical force. In order to verify that the nominal current of 1kA DC as well as the short-circuit current of 8kA DC can be measured, lab tests at the Austrian Institute of Technology are carried out and are already planned for October 2023.

7. LITERATURVERZEICHNIS:

Arar, D. S. (6. 6 2022). *allaboutcircuits*. Von allaboutcircuits: https://www.allaboutcircuits.com/technical-articles/hall-effect-current-sensing-open-loop-and-closed-loop-configurations

Berthou, A. (11. 6 2020). *The Benefits of High-Voltage Direct Current (HVDC) PowerDirect current in the transmission gird*. Von eepower: https://eepower.com/technical-articles/the-difference-that-dc-makes

DKE. (1983). *Hochspannungspürftechnik – Teilentladungsmessung (IEC 38/516/CD:2016).* VDE-Verlag GmbH.

KASAI, A. (12. 5 2023). *akm.* Von https://www.akm.com/content/dam/documents/products/magnetic-sensor/hall-element/in-sb-ultra-high-sensitivity/hw322b

R. Portas, L. C. (10. 12 2022). *www.iconpower.com*. Von iconpower: https://www.iconopower.com/v/lem/coulometry.pdf

TI. (1. 8 2020). *Texas Instruments*. Von https://www.ti.com/lit/ds/symlink/drv411.pdf?ts=1694071479291

Xanthakis, J. (2020). *Electronic Conduction: Classical and Quantum Theory to Nanoelectronic Devices.* CRC Press.



Michael Holzbauer
Dr. techn. Josef ZELISKO Fabrik für Elektrotechnik und Maschinenbau GmbH, Beethovengasse 43-45, 2340 Mödling

Gesundheitstechnik und Medizininformatik an der HTL Spengergasse in Wien (Matura 2016), Bachelor-Studium „Mechatronik und Mikrosystemtechnik“ an der FH Wiener Neustadt 2017 bis 2020, Werkstudent bei Zelisko seit 2020, Mechanics Engineer bei Zelisko seit März 2023; Master-Studium „Mechatronik“ an der FH Wiener Neustadt seit 2020

CONTROL ELECTRONICS FOR ON-CHIP LASER SPECTROSCOPY

Knöbelreiter, L.

Abstract: *Intensified research in silicon photonics has led to the discovery of CMOS compatible materials suitable for visible and near-infrared light, such as e.g. silicon nitride. This opens a door to new integrated optoelectronics, one of which is integrated frequency-stabilised Doppler-free spectroscopy on alkali atoms. In this thesis, first the feasibility of this concept is investigated using spectroscopy on Rubidium (Rb). For this, for each spectroscopy component on-chip solutions are presented, where all components are CMOS compatible and suitable for light exhibiting the D_2-transition wavelength of λ=780 nm. Furthermore, a PI-controller, which stabilises the frequency of a laser to the Rb 85 D_2-transition line, is developed in this thesis. The schematic and PCB layout are built up in Altium Designer and the PCB is then fabricated by milling. Finally, the fabricated PCB is tested with a Doppler-free spectroscopy set-up, where the circuit has successfully locked the laser light onto the D_2-transition wavelength.*

Keywords: *laser stabilisation, PI-controller, on-chip Doppler-free saturation spectroscopy, silicon nitride*

1. INTRODUCTION

Since the size of electronic integrated circuits gradually, but surely reaches its physical limits, optical integrated circuits have come more and more into focus. Using light as a medium of transmission holds many great advantages like extremely high data rates, low power devices and extreme compactness [1]. Until recently, silicon-based optical circuits operating at the telecommunication wavelengths in the infrared were the option of choice, due to the high compatibility with the standardized complementary metal oxide semiconductor (CMOS) processes. However, silicon is only transparent to light with wavelengths above λ > 1.1µm [2], which has limited the developed devices to this wavelength range. Intensified research has led to the discovery of new materials, like silicon nitride or aluminium nitride, exhibit transparency in visible and near infrared light wavelength ranges. These are opening the door to a set of new optical on-chip applications, one of which being an integrated frequency-stabilised Doppler-free spectroscopy.

In order to laser-cool Rubidium (Rb) atoms, a laser must operate at the very precise D2-line transition wavelength of Rubidium 85, which is λ_{D_2}= 780.24 nm. To ensure this, Doppler-free spectroscopy needs to be carried out e.g. using Rubidium 85 atomic vapour, in order to determine the exact transition frequency. After this the laser frequency needs to be matched to the found frequency, which is done by a PI-controller. The controller adjusts the laser frequency depending on the value of an error signal, which is generated from the spectroscopy signal. This means for an entire integrated frequency-stabilised Doppler-free spectroscopy application both parts (the optical spectroscopy circuit and the electrical PI-controller circuit) must be combined. Figure 1 shows a schematic overview of an entire integrated frequency stabilised spectroscopy set-up.

In this work the first important foundation for a completely integrated frequency-stabilised Doppler-free spectroscopy application will be laid. In the first step the theory describing the functionality of a laser system and the working principle of spectroscopy will be covered. After this a thorough feasibility study regarding the integration of the spectroscopy set-up will be carried out. For this an on-chip approach for each individual part of the spectroscopy set-up will be presented. Here the focus will be laid on ensuring that the used materials are CMOS and λ=780 nm compatible. Then all the integrated components will be brought together, to give an idea of what a complete on-chip spectroscopy application could be composed. Next the electrical circuit will be developed and tested. With this a PI-controller for stabilising the laser frequency will be developed. Here a schematic and corresponding layout will be created, built up and tested with different prototypes (breadboard, PCB), where two different versions of the schematic will be created. The PCB will finally be tested with an input signal generated by an actual Doppler-free spectroscopy set-up.
Due to the complete compatibility of all components (electrical controller and optical spectroscopy elements) with the CMOS process this thesis could be the first move towards eventually combining these components into one integrated optoelectronic device (compare Figure 1).

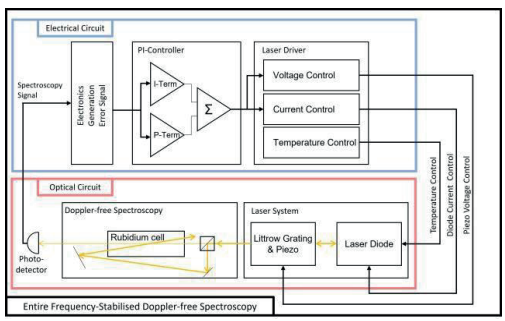


Figure 1 - Schematic overview of an entire integrated frequency stabilised spectroscopy set-up.

2. PROBLEM AND GOAL DEFINITION

The goal of this thesis is to investigate both the electrical and the optical subsystem of an entire integrated frequency-stabilised spectroscopy and with this take the first steps towards the realisation of this application. For this regarding the optical subsystem a feasibility study is carried out, showing if and how an integrated optical spectroscopy set-up can be realised. For the electrical circuit an outdated PI-controller is redeveloped after newly defined requirements. This circuit is then fabricated and tested for the desired functionality.

3. LASER FREQUENCY STABILISATION

The PI-controller was developed to stabilise the frequency of a diode laser system. Such a system provides light produced by a semiconductor laser diode via stimulated emission. Although the laser beam has a very narrow linewidth its frequency still always drifts a bit due to mechanical effects and temperature variations. Hence measures for stabilising the laser

frequency onto the needed transition frequency are applied: passive frequency stabilisation via an optical feedback (Littrow Configuration) and active frequency stabilisation via a PI-controller in combination with Doppler-free absorption spectroscopy for determining the desired frequency.

3.1 Active Stabilisation – Doppler- free Spectroscopy and Laser Locking

For laser cooling the wavelength must fit to a specific atomic transition. Hence the laser wavelength must be measured, compared to the transition wavelength and adjusted depending on the occurring error. This procedure is called active frequency stabilisation or laser locking and is performed via an electric circuit (PI-controller). For this, first the laser light's current frequency must be measured accurately which is done via Doppler-free spectroscopy.

When a laser beam's frequency is resonant with an atomic transition and is sent into atomic vapour a decrease in intensity is measured because the atoms absorb the light. This absorption dip would make it easy to precisely obtain the atoms' transition frequencies. The measured absorption function though is broadened by occurring Doppler-effects to such an extent, that the absorption dip vanishes and the transition frequency cannot be determined anymore. Doppler-free spectroscopy circumvents these effects which allowing the transition frequencies to be determined precisely.

The Doppler-free spectroscopy set-up is schematically depicted in Figure 2. The set-up consists of two light beams propagating through a cell filled with Rb atomic vapour, where the beams propagate in opposite directions. The laser beam is split and redirected so that the pump beam exhibits a significantly higher intensity than the probe beam. The high intensity beam pumps the atoms into an excited state until they saturate. The low intensity beam is the actual spectroscopy signal and is measured by a photo diode.

The measurement of the transition frequencies utilises the compensation of the Doppler broadening by the two beams. The frequency of the base beam scans through a frequency range by controlling the piezo of the Littrow Configuration. Only if the beam's frequency is resonant with an atomic transition both beams excite the same atoms. The pump beam has such a high intensity, that it excites the atoms to saturation. Thus, the probe beam travels through the cloud nearly unaffected as it will not be absorbed by the saturated cloud. Hence its intensity then peaks at the transition frequency because less light gets absorbed. This peak, which is visible in the measured distribution, indicates the precise (Doppler-free) location of the transition frequency.

For laser locking the controller needs an error signal as an input to tune the frequency accordingly. For this the first derivative of the spectroscopy signal is calculated and the resulting signal represents the error signal. This is because the peak indicating the transition frequency has zero gradient at the maximum, hence the zero-line crossing of the derivative lies exactly at the transition frequency. Hence the controller trying to keep the error signal zero results in holding the laser frequency locked onto the desired transition frequency.

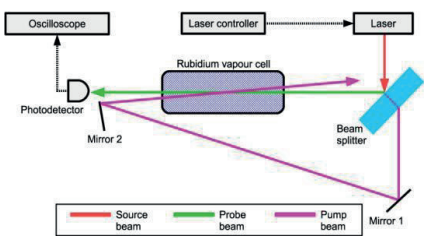


Figure 2 - Schematic depiction of a Doppler-free spectroscopy set-up [3].

4. FEASIBILITY STUDY - OPTICAL SUBSYSTEM

As the idea of this work being to lay the foundations for a complete on-chip frequency-stabilised Doppler-free spectroscopy, the next goal is to carry out a feasibility study to show that spectroscopy (compare Figure 2) performed with Rb can be realised on-chip. Due to intensive research new silicon materials applicable at visible (VIS) and near-infrared (NIR) light have been discovered. With the D_2-line transition frequency lying at λ=780.24 nm these discoveries make an integrated Rb spectroscopy a reachable goal. The dielectric material silicon nitride (Si_3N_4) is a promising candidate as it is CMOS compatible and transparent for the wavelength of interest [2]. In order to realise the spectroscopy optical circuit every on-bench component needed for spectroscopy (mirrors, photodiode, vapour cell, etc.) needs an integrated equivalent. Furthermore, additional components like waveguides or couplers are needed as well. The feasibility study shows that every component already has been in one or another form been realised on-chip. An example shall be given by the integrated version of a beam splitter.

Standard on-bench beam splitters are glass cubes or plates which split an incoming beam into, usually two, sub-beams. In [4] an integrated beam splitter version is presented. It is based on a common directional coupler which consists of two parallel waveguides, which exchange intensity due to evanescent fields formed in the substrate. What amount of light is coupled, depends on the gaping between the waveguides and the coupling length L_C. In [4] to achieve polarisation filtering and lower losses a coupler with a special coupler structure has been implemented. This coupler can be seen in Figure 3.

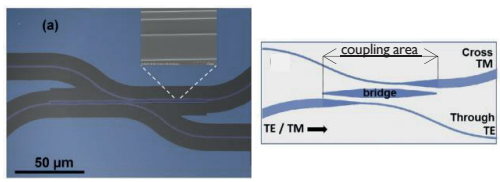


Figure 3 – Schematic and fabricated picture of the integrated beam splitter [4].

Finally, all the integrated components can be put together to give an idea of what an entire on-chip spectroscopy circuit could look like. A schematic example of the optical circuit can be seen in Figure 4.

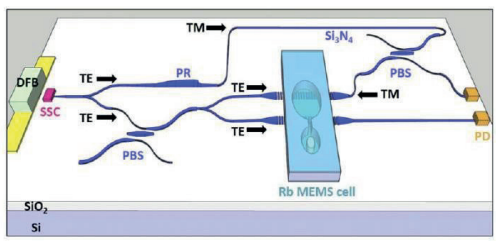


Figure 4 - Schematical depiction of an entire integrated spectroscopy set-up [4].

5. ELECTRICAL SUBSYSTEM – LASER LOCKING ELECTRONICS

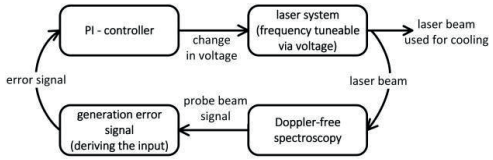


Figure 5 - Schematic depiction of the entire control loop.

The ultimate goal is a frequency stabilised on-chip spectroscopy set-up providing laser light, which frequency is already fixed to the very precise D_2-transition in Rb, so it can be directly used for atom cooling or in an atomic clock, see Figure 5. The needed PI-controller was developed in this work, where an existing design of an analogue PI-controller was altered and redeveloped after redefined requirements. A block diagram of the entire developed controller circuit is depicted in Figure 6.

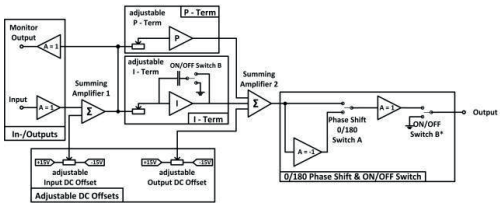


Figure 6 - Block diagram of the entire controller circuit.

The circuit was built up with classic operational amplifier circuits like buffer, integrator, summing circuit, etc. First the schematic was built with the design tool *Altium Designer*. As shown in Figure 6 the circuit can be divided into sub-blocks. Each one's theoretical layout and actual schematic realisation is presented and described in the thesis.

In the next step the first prototype was built by breadboarding the schematic. For this taking the requirements under consideration, the *LM741* from Texas Instruments[1] was the op-amp of choice. This prototype was tested within a simplified testing environment to first verify if the developed circuit exhibits the expected behaviour and with that fulfils all the requirements.

Since the breadboard showed a satisfying behaviour a PCB layout was created in *Altium Designer* after the schematics and subsequently the according PCB was fabricated via milling. The finished PCB can be seen in Figure 7, where all in- and outputs are labelled. The PCB was first tested again with the simplified set-up to check if its behaviour was as desired, where it showed similarly satisfying behaviour as the breadboard prototype.

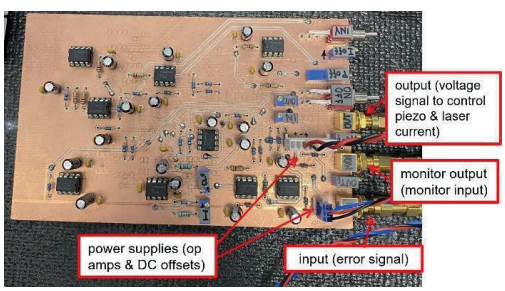


Figure 7 - Picture of the fabricated PCB.

[1] https://www.ti.com

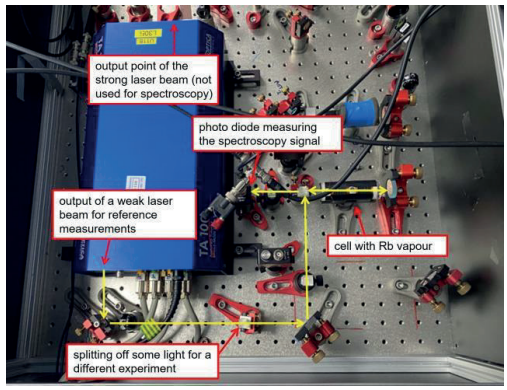


Figure 8 - Spectroscopy set up on an optical bench.

After this check a test in a lab environment with a real spectroscopy signal as an input was carried out. The Doppler-free spectroscopy built up on an optical bench used for this can be seen in Figure 8.

6. RESULTS

The test in the lab showed that the PCB successfully locked the laser frequency onto the desired Rb transition frequency. Furthermore all new features explicitly added to create an easier and more effortless locking process than with the outdated circuit version functioned as desired. Figure 9 shows the entire Rb absorption valley, where the transmission peaks represent the hyperfine level transitions. Alongside the spectroscopy signal (purple) the corresponding derivative (green) functioning as the error signal, can be seen. The controller locks onto one of the purple peaks, by keeping the corresponding error signal equal to zero.

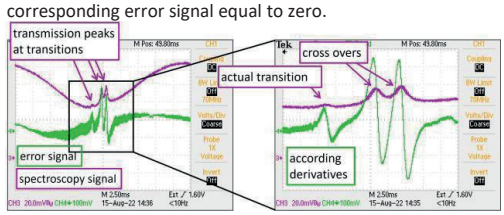


Figure 9 – Depiction (right) and further zoom (left) of the spectroscopy signal (purple) showing one of the Rubidium 85 absorption valleys with their hyperfine structures and the error signal (green) representing the corresponding derivative.

In Figure 10 a successful locking process is shown. Here the laser frequency was locked onto the (rather small) peak of an atomic transition (most left peak in Figure 9). Figure 10 (left) shows the situation before locking, whereas (right) after locking and it can be seen, that here the error signal is kept at zero, as expected.

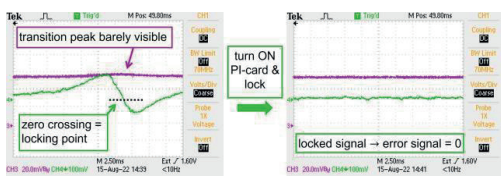


Figure 10 - Successful locking process of locking the laser onto the transition peak of an atomic transition.

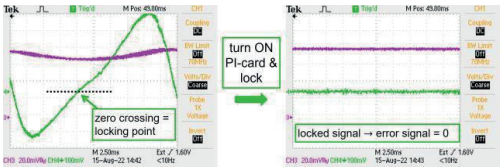


Figure 11 - Successful locking with the error signal crossing the zero-line from the lower half plane.

One important requirement was that the controller must be able to lock either way if the error signal crosses the zero-line from either the upper or the lower half plane. Figure 11 shows that the controller also locked successfully when the error signal crossed the zero-line from below.

There were two versions of the schematic and PCB layout created. The first version was supplied by common power supplies. By adding a corresponding connector and voltage regulators the second version will in the future be supplied by a lab rack.

6. SUMMARY AND OUTLOOK

In this work the first foundation for an integrated frequency stabilised Doppler-free spectroscopy was laid. For this first a feasibility study regarding the integration of the optical components required for an on-chip realised spectroscopy set-up was carried out. With this for each component a possible integrated realisation approach was presented, where the focus was laid on ensuring CMOS and λ=780 nm wavelength compatibility. After this the electrical circuit, for stabilising the laser frequency, came into focus. For this a PI-controller, which controls and stabilises the laser frequency onto the Rubidium 85 D_2-transition frequency was developed after newly defined requirements. An according electrical schematic and PCB layout was built up with the design tool *Altium Designer*. The circuit underwent multiple tests, with different prototypes and different testing environments. It was verified that the circuit showed the desired behaviour (integrates, amplifies, etc. correctly) and it was checked that the PCB locks the laser frequency successfully. Both points were fulfilled, hence the developed PI-controller fulfils all defined requirements.

For the ultimate goal of combining the two subsystems to an entire integrated frequency-stabilised Doppler-free spectroscopy all the materials of the individual optical parts would need to be matched to each other and tested thoroughly, so that that the optical circuit could be put together. The electrical circuit would need to be built up integrated circuit and then again the two sub-circuits (optical & electrical) would need to be matched and connected to each other.

7. REFERENCES

[1] X. Mu, S. Wu, L. Cheng, and H. Fu, “Edge couplers in silicon photonic integrated circuits: A review,” *Applied Sciences*, vol. 10, no. 4, 2020.

[2] A. Z. Subramanian, P. Neutens, A. Dhakal, et al., “Low-Loss Singlemode PECVD Silicon Nitride Photonic Wire Waveguides for 532–900 nm Wavelength Window Fabricated Within a CMOS Pilot Line,” *IEEE Photonics Journal*, vol. 5, no. 6, pp. 2 202 809–2 202 809, 2013.

[3] A. Wyngaard, “Saturated absorption spectroscopy of rubidium and feedback control of laser frequency for doppler cooling,” M.S. thesis, Cape Peninsula University of Technology, February 2018.

[4] K. Gallacher, P. F. Griffin, E. Riis, M. Sorel, and D. J. Paul, “Silicon nitride waveguide polarization rotator and polarization beam splitter for chip-scale atomic systems,” *APL Photonics*, vol. 7, no. 4, p. 046 101, 2022.



Lisa Knöbelreiter, BSc

SUPPORTING INSTITUTIONS

University of Nottingham & Fachhochschule Wiener Neustadt (cooperation)
Uni of Nottingham: University Park, Nottingham, NG7 2RD, United Kingdom
FH Wr. Neustadt: Johannes-Gutenberg-Straße 3, 2700 Wiener Neustadt, Austria

SHORT CV

Education:
2012 – 2017 A-levels completed at the Higher Institute for Economic Professions Wiener Neustadt
2017 – 2020 Bachelor programme: Mechatronics & Microsystems engineering at FH Wiener Neustadt
2020 – 2023 Master programme: Mechatronics at FH Wiener Neustadt
Working experience:
2018 & 2019 Summer Internships at Magna Powertrain GmbH & Schiebel Aircraft GmbH
01/2020 – 04/2020 Bachelor project incl. writing the Bachelor thesis at Magna Powertrain GmbH
10/2021 – 01/2022 Working student in System Design Department at Magna Powertrain GmbH
03/2022 – 09/2022 Master project at University of Nottingham
09/2022 – now Master project and finishing Master thesis at FH Wiener Neustadt

THEMA MASTERARBEIT
ENTWICKLUNG EINES PRÜFSTANDES FÜR HOCHLEISTUNGSPROJEKTOREN

Kraft, J.

Kurzfassung: *Die Masterarbeit verfolgt das Ziel, einen Prüfstand zur präzisen Kalibrierung von Hochleistungs-DLP-Projektoren im Kontext der Stereolithographie zu konzipieren und zu realisieren. Angesichts der essenziellen Bedeutung exakter Kalibrierungsverfahren für die Fertigung hochqualitativer Bauteile und der aktuellen Limitationen hinsichtlich Zeit- und Fehleranfälligkeit der etablierten Methoden, gewinnt dieser Forschungsbereich signifikante Relevanz. Ein primäres Augenmerk liegt dabei auf der Adaption des Prüfstands auf das Hammer Lab35-System von Incus.*

Die Arbeit postuliert die Implementierung einer Softwarekomponente, welche optische Messungen mittels einer neuartigen Methodik durchführt, welche Python, OpenCV und Bildverarbeitungstechniken einschließt. Hierdurch wird die Präzision der Messprozesse gesteigert und potenzielle subjektiven Unschärfen minimiert. Die entwickelte Methodik quantifiziert Abstände zwischen projiziertem Licht des DLP-Projektors und dem Kalibrierungsplatte. Die resultierende Möglichkeit einer präzisen Charakterisierung der Projektoren ermöglicht eine umfassende Vorabkalibrierung vor der Einbindung in das Gesamtsystem. Die Effekte sind eine erhebliche Reduktion des Kalibrierungsaufwands sowie Kosteneinsparungen durch präventive Fehlervermeidung.

Schlüsselwörter: *Prüfstand, DLP-Projektoren, Kalibrierung, Messung, LMM, Stereolithographie, Python, Bildverarbeitung*

1. EINLEITUNG

Die Stereolithographie, als etabliertes additiv-fertigendes Verfahren, nutzt DLP-Hochleistungsprojektoren als entscheidende Lichtquellen für die strukturgebende Polymerisation flüssiger Photopolymere [1]. Die Präzision und Leistung dieser Projektoren sind maßgeblich für die Qualität der erzeugten Bauteile. Bedauerlicherweise sind existierende Kalibrierungsmethoden zeitaufwändig und anfällig für Fehler. In dieser Hinsicht manifestiert sich die Dringlichkeit, einen effektiven Prüfstand zu konzeptionieren, der durch Einsatz von Python, OpenCV und Bildverarbeitung eine automatisierte, präzise Beurteilung der Projektorenqualität vor der Systemintegration ermöglicht. Durch diese Entwicklungsarbeit wird beabsichtigt, den zeitraubenden und fehlerbehafteten Prozess der manuellen Kalibrierung zu überwinden. Die gewonnenen Resultate manifestieren, dass der entwickelte Prüfstand Kalibrierungszeit einspart, Kosten minimiert und die Qualität der additiv gefertigten Bauteile steigert. Dieses Unterfangen hat weitreichende Anwendungseinsätze für vielfältige Industriezweige, wie die Medizintechnik, Luftfahrt und den Maschinenbau, indem die Effizienz erhöht und die Produktionszeit verkürzt, werden können. Die gegenwärtige Arbeit steuert zum wissenschaftlichen Bereich der Projektorkalibrierung für die Stereolithographie bei und stärkt die Applikation dieser Technologie in verschiedenen Domänen.

2. PROBLEM- UND AUFGABENSTELLUNG

Im Kontext der Stereolithographie werden DLP-Hochleistungsprojektoren zur schichtweisen Herstellung komplexer Bauteile aus flüssigem Photopolymer eingesetzt. Derzeit existieren unterschiedliche wissenschaftliche Arbeiten, die sich unter anderem mit der Intensitätsmessung von Projektoren und der Kalibrierung sowie Pixelschärfe beschäftigt haben, siehe Tabelle 1. Jedoch untersuchen diese Arbeiten nicht im notwendigen Ausmaß das Optimierungspotenzial- sowie Automatisierungspotenzial von Kalibrierungsverfahren. Die derzeitige Kalibrierung von Projektoren ist zeitaufwendig und fehleranfällig, da sie von subjektiven Einschätzungen abhängt. Das Hauptproblem liegt in der mangelnden Präzision und Automatisierung der bestehenden Kalibrierungsmethoden. Die vorliegende Forschungsarbeit zielt darauf ab, ein neuartiges Verfahren für die Kalibrierung von DLP-Hochleistungsprojektoren in der Stereolithographie zu entwickeln. Dabei sollen automatisierte und präzise Kalibrierungsmethoden im Fokus stehen, um menschliche Fehler zu minimieren. Die Kernfragen sind, wie sich die Genauigkeit und Effizienz der Kalibrierung verbessern lassen, welche Herausforderungen dabei auftreten und wie moderne computergestützte Technologien dabei eingesetzt werden können. Die Forschung zielt darauf ab, die Qualität der gedruckten Bauteile zu optimieren und die Kalibrierungszeit zu reduzieren, um die Leistung der additiven Fertigung zu steigern.

Tabelle 1: Liste von Veröffentlichungen zur Projektor-Kalibrierung

Nr	Titel	Autoren	Jahr
1	Accurate calibration for a camera-projector measurement system based on structured light projection [2]	Xiaobo Chen, Juntong Xi, Ye Jin, Jin Sun	2008
2	Accurate and Simple Calibration of DLP Projector Systems [3]	Jakob Wilm, Oline Vinter Olesen, Rasmus Larsen	2014
3	An Accurate Projector Calibration Method Based on Polynomial Distortion Representation [4]	Miao Liu, Changku Sun, Shujun Huang, Zonghua Zhang	2015
4	A New Calibration Method of the Projector in Structured Light Measurement Technology [5]	Zhen Lei, Xiaojun Liu, Wenlong Lu, Zili Lei, Liangzhou Chen, Liping Zhou	2015
5	Analysis of different camera calibration methods on a camera-projector measuring system [6]	Julieta Tiscareno, José A. Albajez, Jorge Santolaria	2019

3. MATERIALIEN UND METHODEN

Die angewandte Methodik greift auf einen systematischen Rahmen zurück, der sich an den Prinzipien der VDI-2222 Richtlinie zur Produktentwicklung orientiert [7]. Dieser methodische Rahmen bietet eine strukturierte Herangehensweise, die sich aus mehreren aufeinanderfolgenden Schritten zusammensetzt: Analyse, Konzeption, Entwurf und Ausarbeitung.

In der Analysephase liegt der Fokus auf dem Verständnis der gegebenen Anforderungen und Rahmenbedingungen. Hierbei werden alle relevanten Informationen gesammelt, um eine solide Grundlage für die kommenden Schritte zu schaffen.

Die Konzeptionsphase nimmt die gesammelten Anforderungen und erarbeiteten Informationen und setzt sie in ein umfassendes Konzept um. Dabei steht die Entwicklung eines detaillierten Konzepts im Mittelpunkt, das die Anforderungen bestmöglich erfüllt. Diese Phase beinhaltet die Erstellung eines morphologischen Kastens. Ein morphologischer Kasten ist ein strukturiertes Kreativitätstool, das verschiedene Lösungsvarianten durch die Kombination verschiedener Parameter generiert. In diesem Kontext werden Lösungsvarianten für einzelne Funktionen wie Rahmen, Antrieb, Führung und Schutz erzeugt und miteinander kombiniert.

Um die vielversprechendsten Lösungswege auszuwählen, kommen Instrumente wie die Auswahlliste und die Nutzwertanalyse zum Einsatz. Die Auswahlliste hilft dabei, die generierten Lösungsvarianten zu erfassen und auf übersichtliche Weise darzustellen. Die Nutzwertanalyse wiederum bietet eine strukturierte Methode, um die verschiedenen Lösungsvarianten anhand vorher festgelegter Kriterien zu bewerten. Dadurch können objektive Entscheidungen getroffen werden, welche Lösungswege die größte Erfolgswahrscheinlichkeit aufweisen.

Die Erarbeitungsphase schließt sich an die Konzeptionsphase an und dient dazu, das ausgearbeitete Konzept in einen detaillierten Entwurf zu überführen. Abbildung 1 illustriert, wie dies mittels der Software Inspire erfolgt. Hier fließen Aspekte wie technische Realisierbarkeit, Materialauswahl und ergonomische Gesichtspunkte mit ein.

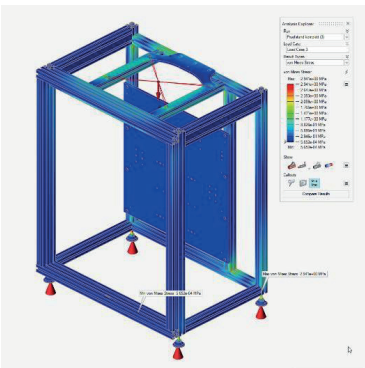


Abb. 1: Berechnung der von-Mises-Spannung des Prüfstandes

4. PRAKTISCHE DURCHFÜHRUNG

Der Implementierungsprozess des Prüfstands erfolgt in mehreren Schritten, um sicherzustellen, dass alle Komponenten ordnungsgemäß integriert werden und der Prüfstand den Anforderungen entspricht. Dieser Prozess beinhaltet die sorgfältige Montage und Konfiguration der Hardwarekomponenten, einschließlich des Gestells, der UV-Schutzvorrichtung, der Positioniereinheit, der Aufnahmeeinheit für den DLP-Projektor und der Tür.

Beim Aufbau des Prüfstand-Gestells werden ausgewählte Konstruktionsprofile verwendet, um ein stabiles und belastbares Rahmenwerk zu schaffen. Abbildung 2 dient zur Visualisierung des Aufbaus des Prüfstandes. Diese Profile werden präzise ausgerichtet und mittels Nutensteinen und Schrauben sicher befestigt. Durch FDM-gedruckte T-Nut-Adapter werden HPL-Platten in die Profile eingeschoben, um einen umfassenden UV-Schutz für den Prüfstand zu gewährleisten.

Die Positioniereinheit, welche die präzise vertikale Bewegung und Positionierung der Z-Achse ermöglicht, wird durch einen Nema17 Hybridschrittmotor und eine Federstegkupplung realisiert. Die Lineareinheit wird an der Rückplatte montiert, wobei Passstifte eine exakte Ausrichtung ermöglichen.

Die Aufnahmeeinheit für den DLP-Projektor besteht aus präzisionsgefertigten Aluminiumplatten, die eine sichere und akkurate Ausrichtung des Projektors gewährleisten. Nivellierschrauben erlauben die präzise Einstellung der Projektorfläche. Die UV-Schutz-Tür besteht aus einer Acrylglasplatte mit UV-Filterfolie, montiert mittels FDM-gedrucktem T-Nut-Adapter, ergonomischem Griff und Sicherheitstürschalter.

Die sorgfältige Montage, Ausrichtung und Prüfung der Hardware sind unerlässlich, um genaue und zuverlässige Messungen sicherzustellen. Dieser Prozess erfordert präzise Planung, Montage und mögliche Feinabstimmungen, um sicherzustellen, dass der Prüfstand den Anforderungen gerecht wird.

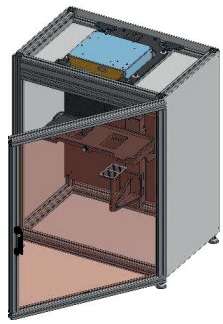


Abb. 2: Aufbau des Prüfstandes

Ergänzend zur beschriebenen Durchführung der Hardwarekomponente erfolgt die Softwarekomponente. Die Softwareentwicklung ist ein essentieller Bestandteil der Prüfstands-Implementierung und ermöglicht die Steuerung, Automatisierung sowie Erfassung, Verarbeitung und Darstellung von Messdaten. Hierbei spielt die Software eine zentrale Rolle bei Präzision, Wiederholbarkeit und Effizienz optischer Messungen. Für die Entwicklung der Software wird Python in Verbindung mit der Bildverarbeitungsbibliothek OpenCV genutzt.

Die entwickelte Software besteht aus eng miteinander interagierenden Modulen, welche die Prüfstandsfunktionalität sicherstellen. Wesentliche Module umfassen:

1. Bildaufnahme: Die Software erfasst Bilder über Webcam oder USB-Kamera und erlaubt die Anpassung von Kameraeinstellungen wie Quelle, Bildbreite und -höhe für optimale Aufnahmebedingungen.
2. Vorverarbeitung der Bilder: Aufgenommene Bilder werden vorbereitet, indem ein Template-Bild für das Matching verwendet wird.
3. Bildanalyse (KO-System): Die Software nutzt Bildverarbeitungstechniken, um Linien im KO-System-Bild zu extrahieren, unter anderem mittels der Hough-Transformation zur Erkennung und Klassifizierung von Linien.

4. Bildanalyse (Schnittpunkt): Ähnlich wie im KO-System erkennt die Software Linien, die Schnittpunkte bilden, und markiert sie entsprechend.
5. Ausgabe und Speicherung: Das bearbeitete Bild mit Linien, Schnittpunkten und Template wird angezeigt und kann unter einem eindeutigen Dateinamen gespeichert werden.
6. Erweiterbarkeit: Die Software bietet Raum für zukünftige Erweiterungen, wie die Integration weiterer Bildverarbeitungsalgorithmen oder die Implementierung einer Benutzeroberfläche zur interaktiven Steuerung.

Die Software läuft auf einem Laptop, welcher über Schnittstellen wie USB und Ethernet mit dem Prüfstand kommuniziert. Dies ermöglicht die effiziente Steuerung, Durchführung von Messungen und Darstellung der Ergebnisse. Die kontinuierliche Weiterentwicklung der Software eröffnet die Möglichkeit zur nahtlosen Integration zukünftiger Anforderungen und Verbesserungen. Insgesamt trägt die entwickelte Software zur Genauigkeit, Effizienz und Benutzerfreundlichkeit des Prüfstands bei, indem sie präzise Steuerung der Hardwarekomponenten, Bildaufnahme, Analyse sowie Ausgabe und Speicherung der Ergebnisse ermöglicht.

5. ERGEBNISSE

Die wissenschaftliche Arbeit präsentiert eine umfassende Analyse der Implementierung der Projektorkalibrierung auf dem Prüfstand. Der Kalibrierungsprozess wurde systematisch entwickelt und eingehend getestet, um präzise Ergebnisse zu erzielen.

Der Evaluierungsabschnitt gliedert sich in zwei Hauptkomponenten: die optische Vermessung und die Intensitätsmessung.

In Bezug auf die optische Vermessung wurde ein detaillierter Aufbau beschrieben, der die präzise Ausrichtung des Projektors und der Kamera ermöglicht. Die automatisierte Bildverarbeitung mittels Python-Skript trägt dazu bei, die Messungen effizient und genau durchzuführen. Das Prozedere des Python-Skripts ist durch Abbildung 3 ersichtlich. Dies wurde durch die Analyse der Messergebnisse bekräftigt, die eine hohe Genauigkeit bei der Bestimmung der Abstände zwischen dem Koordinatenursprung und den beleuchteten Punkten auf der Kalibrierglasplatte zeigten.

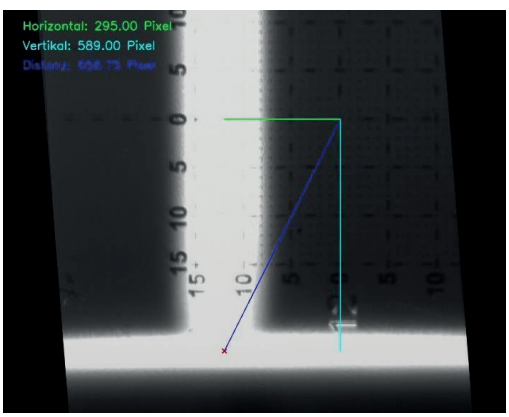


Abb. 3: Ergebnis der optischen Vermessung mit Python-Skript

Die Intensitätsmessung wurde ebenfalls detailliert erläutert, wobei der Aufbau und der Ablauf beschrieben wurden. Hierbei wurde eine systematische Vorgehensweise angewandt, um die Intensität der belichteten Kreise präzise zu messen. Die iterative Anpassung der Intensität auf Basis der gemessenen Werte führte zu einer hohen Übereinstimmung mit den Zielintensitäten.

Die Analyse der Messergebnisse verdeutlicht, dass der entwickelte Kalibrierungsprozess die gestellten Anforderungen erfüllt und präzise Kalibrierungsergebnisse liefert. Die erreichte Genauigkeit der optischen Vermessung und der Intensitätsmessung erfüllt die Qualitätsstandards für den 3D-Druck. Insbesondere die automatisierte Bildverarbeitung hat sich als leistungsstarkes Instrument erwiesen, um genaue Messungen durchzuführen. In Abbildung 4 wurden die Ergebnisse von einzelnen Mitarbeitern gegenüber dem automatisierten Verfahren mittels Python-Skript gegenübergestellt. Die Genauigkeit hat sich ersichtlich verbessert.

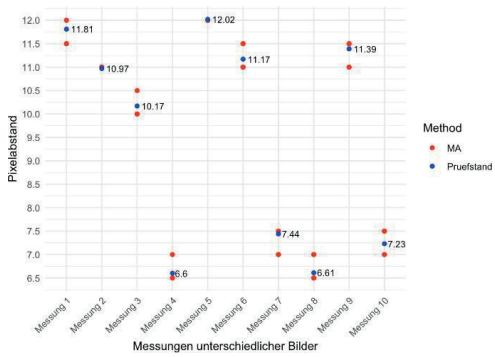


Abb. 4: Gegenüberstellung der Messungen von Mitarbeitern und Prüfstand mit Python-Skript

Die Zeitanalyse zeigt deutlich, dass der optimierte Kalibrierungsprozess die Kalibrierungszeit erheblich verkürzt, ohne die Genauigkeit zu beeinträchtigen. Dies stellt eine wesentliche Verbesserung gegenüber dem vorherigen Verfahren dar und unterstreicht die Effizienz des Prüfstands.

Die Entwicklung und Implementierung des optimierten Kalibrierungsprozesses bietet somit eine solide Grundlage für präzise und effiziente Projektorkalibrierungen im Bereich des 3D-Drucks. Die Kombination aus automatisierter Bildverarbeitung, sorgfältiger Vorbereitung und präziser Messung gewährleistet eine zuverlässige Kalibrierung und führt letztendlich zu hochwertigen Ergebnissen in der 3D-Drucktechnologie.

6. ZUSAMMENFASSUNG UND AUSBLICK

Die Entwicklung eines Prüfstandes adressierte intensiv die Problematik der präzisen Kalibrierung von DLP-Projektoren im Rahmen der Stereolithographie. Die bisher manuell durchgeführte Kalibrierungsprozedur erwies sich als zeitaufwändig und anfällig für menschliche Fehler. Die präsentierte Methodik integriert fortgeschrittene Bildverarbeitungstechniken, um genaue Evaluierungen der Projektionsqualität zu ermöglichen und somit den Gesamtprozess der Kalibrierung zu optimieren.

Die resultierenden Erkenntnisse aus den durchgeführten Messungen und Analysen unterstreichen die Wirksamkeit des skizzierten Vorgehens. Die automatisierte Bildverarbeitung und die exakte Berechnung der Abstände zwischen dem projizierten

Licht eines DLP-Projektors und der Oberfläche des Kalibrierungsmediums generierten präzise und reproduzierbare Kalibrierungsergebnisse. Die Implementierung des entwickelten Prüfstandes erwies sich als äußerst effizient und ermöglichte signifikante Zeitersparnisse bei der Durchführung der Projektorkalibrierung.

In Bezug auf potenzielle Anwendungsgebiete zeichnet sich ab, dass der etablierte Prüfstand eine breite Bandbreite von Anwendungsbereichen abdeckt, angefangen bei der Wareneingangskontrolle von DLP-Projektoren bis hin zur präzisen Feinjustierung solcher Projektoren und der automatisierten Überwachung von Projektionssystemen. Hierdurch wird eine substanzielle Qualitätsverbesserung, Fehleridentifikation sowie Prozessoptimierung ermöglicht.

Ausblickend auf zukünftige Entwicklungen und Verfeinerungen, eröffnen sich vielfältige Erweiterungs- und Optimierungsmöglichkeiten. Die Integration eines Spektrometers könnte beispielsweise die Qualitätsevaluierung weiter verfeinern, während motorisierte Mechanismen die Automatisierung des Prüfsystems vorantreiben könnten. Des Weiteren könnten die Implementierung von nutzerfreundlichen grafischen Benutzerschnittstellen und automatischen Fokusfunktionen zur Steigerung der Anwenderfreundlichkeit und Messpräzision beitragen.

Zusammenfassend trägt die gegenwärtige Untersuchung zur Optimierung der Stereolithographietechnologie bei, indem sie einen leistungsfähigen Prüfstand zur Kalibrierung von DLP-Projektoren entwirft. Mittels dieser Neuerungen kann die Effizienz und Verlässlichkeit des Kalibrierungsprozesses spürbar gesteigert werden. Die entwickelten Methoden und Resultate bieten vielseitige Anwendungsmöglichkeiten in verschiedenen Industriezweigen und Forschungsdisziplinen und ermöglichen die Produktion hochwertiger Produkte auf effiziente und zuverlässige Art und Weise.

7. LITERATURVERZEICHNIS

[1] Chen, X., Xi, J., Jin, Y., & J. Sun 2009. *Accurate calibration for a camera-projector measurement system based on structured light projection*, Optics and Lasers in Engineering, vol. 47, no. 3-4, pp. 310-319

[2] Hull, C, 1984. *Apparatus for production of three-dimensional objects by stereolithography*, U.S. Patent US4 575 330A

[3] Lei, Z., Liu, X., Lu, W., Lei, Z., Chen, L. & Zhou L., 2015. *A New Calibration Method of the Projector in Structured Light Measurement Technology*, Procedia CIRP, vol. 27, pp. 303-308

[4] Liu, M., Sun, C., Huang, S. & Zhang, Z., 2015. *An Accurate Projector Calibration Method Based on Polynomial Distortion Representation*, Sensors, vol. 15, no. 10, pp. 26 567-26 582

[5] Tiscareno, J., Albajez, J. & Santolaria, J., 2019. *Analysis of different camera calibration methods on a camera-projector measuring system*, Procedia Manufacturing, vol. 41, pp. 539-546

[6] Verein Deutscher Ingenieure, 1997. *VDI 2222-Konstruktionsmethodik Methodisches Entwickeln von Lösungsprinzpien* Blatt 2

[7] Wilm, J., Olesen, O., & Larsen R., 2014. *Accurate and simple calibration of DLP projector systems*, SPIE MOEMS-MEMS, M. R. Douglass, P. S. King, and B. L. Lee, Eds., San Francisco, California, United States, p. 897909



Johannes Kraft

Incus, Christine-Touaillon-Straße 11, Technologiezentrum Seestadt 2, 1220 Wien

Bachelor-Studium „Maschinenbau“ an der Technische Hochschule Mittelhessen,
Master-Studium „Mechatronik“ an der FH Wiener Neustadt seit 2021

EINFLUSS VON INNENSTRUKTUREN UND STÜTZSTRUKTUR-PARAMETERN AUF DIE HIRTISIER®BEHANDLUNG VON LBM-BAUTEILEN

Kokic, A.

Kurzfassung: *Eine neuartige Methode zur Oberflächenbehandlung von AM-Bauteilen ist die Technologie des Hirtisierens®. Diese Methode weist mehrere Vorteile gegenüber herkömmlichen Nachbehandlungsmethoden auf und bedarf weiterer Optimierung. Mit generativen Fertigungsverfahren hergestellte Bauteile benötigen häufig Stützstrukturen, die nach der Fertigung entfernt werden müssen. Es soll untersucht werden, welche Stützstruktur-Parameter für einen erfolgreichen Hirtisier®prozess entscheidend sind. In bisherigen Versuchen wurden Besonderheiten dieser Methode beobachtet, die einer weiteren Analyse bedürfen. So soll ermittelt werden, ob Bauteile mit einem geringeren Verhältnis von Innenfläche zu Gesamtoberfläche einen Einfluss auf den mittleren Abtrag haben. Wie die in den Elektrolyten eingetauchten Bauteile am schnellsten mit Strom kontaktiert werden können, wird ebenfalls Gegenstand dieser Arbeit sein.*

Schlüsselwörter: *Hirtisieren®, Oberflächenbehandlung, Stützstrukturen, generative Fertigungsverfahren, Klemmvorrichtung*

1. EINLEITUNG

Die gegenständliche Arbeit behandelt eine weiterentwickelte Variante des elektrochemischen Polierens, das sog. Hirtisieren®. Während das klassische Elektropolieren entweder gepulst oder ungepulst arbeitet, handelt es sich dabei um eine neuentwickelte, automatisierte Variante des Elektropolierens mit speziellen Pulsformen, die sich von der Pulsformen des klassischen Elektropolierens unterscheiden [1]. Der Schwerpunkt der Untersuchungen soll die Erweiterung des Knowhows zur Wechselwirkung von 3D-Druck- und Hirtisier®parametern sein. Das primäre Ziel ist es, einen Beitrag zur Optimierung der Hirtisier®behandlung von LBM-Bauteilen zu leisten. Eine Weiterentwicklung des Finishing Moduls H3000, eine Weiterentwicklung von Elektrolyten für die Hirtisier®behandlung sowie eine Erprobung der in dieser Arbeit entwickelten Klemmvorrichtung zählen zu Nicht-Zielen dieser Arbeit.

2. PROBLEM- UND AUFGABENSTELLUNG

Um eine erfolgreiche Entfernung der Stützstrukturen zu gewährleisten, müssen sie sich im Elektrolyten gut auflösen können. Trotzdem müssen sie ihre eigentliche Funktion erfüllen, d.h. sie sollten eine ausreichende Festigkeit aufweisen. Daher müssen die Parameter der Stützstruktur so gewählt werden, dass ein Kompromiss zwischen ausreichender Festigkeit und guter Entfernbarkeit mittels Hirtisieren® gefunden wird. Es soll systematisch untersucht werden, wie sich diese Parameteränderungen auf die mechanische Festigkeit der Strukturen auswirken. Außerdem wurden Besonderheiten des Hirtisieren®s beobachtet, die einer weiteren Analyse bedürfen. So soll ermittelt werden, ob Bauteile mit einem geringeren Verhältnis von Innenfläche zu Gesamtoberfläche einen Einfluss auf den mittleren Abtrag haben. Wie die in den Elektrolyten eingetauchten Bauteile am schnellsten mit Strom kontaktiert werden können, wird ebenfalls Gegenstand dieser Arbeit sein.

3. MATERIALIEN UND METHODEN

Laserstrahlschmelzen

Für die Untersuchungen in dieser Arbeit wurden die Bauteile aus Aluminiumlegierung AlSi10Mg mit dem LBM-System M400 (EOS) gebaut. Das LBM-Verfahren wird auch als selektives Laserschmelzen (SLM) oder DMLS-Verfahren (=„Direct Metal Laser Sintering“) bezeichnet. Das Prinzip dieses Verfahrens beruht auf dem Hinzufügen von Materialschichten. Der Prozess läuft in drei Schritten ab. Im ersten Schritt wird eine Metallpulverschicht auf der Bauplattform aufgetragen [2]. Diese Schicht wird mit Hilfe eines Laserstrahls dort aufgeschmolzen, wo das Bauteil entstehen soll. Nachdem eine Schicht (= Layer) aufgeschmolzen ist, wird sie durch Abkühlen zu einem festen Körper erstarrt [3].

Hirtisieren®

Die Finishing-Module der H-Serie von Hirtenberger Engineered Surfaces (HES) sind die einzigen Anlagen, die das Verfahren Hirtisieren® durchführen können. Das H3000 Finishing Modul wurde speziell für Kleinserienfertigung und für Forschungs- und Entwicklungsaufgaben entwickelt. Die Anlage dient zum gleichzeitigen Abtragen, Polieren und Reinigen von Bauteilen mit dem Verfahren des Hirtisieren®s. Die Maschine eignet sich für alle 3D gedruckten Metalle und Legierungen. Es wird auch für eine gleichzeitige Entfernung von Stützstrukturen verwendet [4]. Das Hirtisier®prozess erfolgt in mehreren Schritten. Bevor der Prozess beginnt, wird das Bauteil fest auf einer Warenschiene fixiert. Diese wird in den Eingabebereich eingelegt und das gewünschte Programm ausgewählt. Der Warenträger wird automatisch vom Roboter übernommen und in das erste Becken eingetaucht. Über einen Gleichrichter wird eine Spannung nach einem entsprechenden Pulsmuster angelegt. Je nach Programm wird der Warenträger im nächsten Bearbeitungsschritt in ein anderes Becken getaucht, in dem der Hirtisier®prozess wiederholt wird. Zwischen den einzelnen Hirtisier®-Prozessschritten kann das Bauteil gespült werden. Nach Abschluss aller Bearbeitungsschritte wird das Bauteil in einem Waschbecken gereinigt und anschließend getrocknet [5]. Die Bauteile sollen die Abmessungen von 300 x 200 x 300 mm (L x B x H) nicht überschreiten und müssen mindestens 100 mm unter den Warenträger liegen. Befinden sich die Bauteile etwa 1-2 Zentimeter unterhalb der Elektrolytoberfläche, kann es zu unerwünschten Reaktionen auf der Oberfläche kommen, z. B. zu Fleckenbildung. Um die gewünschte Oberflächengüte zu erreichen, muss das Bauteil in einem Abstand von 2 cm von der Elektrolytoberfläche platziert werden [6].

Aluminiumlegierung AlSi10Mg

Eine der am häufigsten verwendeten Aluminiumgusslegierungen, die sich für das Gießen in verschiedenen Verfahren eignet, ist die Legierung AlSi10Mg.

Die mechanischen Eigenschaften, die Gießbarkeit und die Korrosionsbeständigkeit sind bei dieser Legierung optimal aufeinander abgestimmt. Deshalb wird sie häufig für hochbeanspruchte Bauteile verwendet. Die aus AlSi10Mg hergestellten Bauteile haben unter anderem gute thermische Eigenschaften und eignen sich für verschiedene Nachbearbeitungsverfahren wie Schweißen, Sandstrahlen und Polieren [7].

Messsystem

Die taktilen Messverfahren, auch als berührende Messverfahren bezeichnet, beruhen auf dem Prinzip des Antastens einer Messfläche. Ein Taster wird mit einer speziell geformten Tastnadel über die annähernd idealen Bezugsfläche gezogen. Je nach der zurückgelegten Strecke wird die Auslenkung der Tastnadel gemessen. Typischer Zielwert ist der arithmetische Mittelwert Rauheit Ra [8]. Unter tragbaren Koordinatenmesssystemen gehören 3D-Messarmen. Diese Messsysteme ermöglichen es, die Position eines Messtasters (Sonde) im dreidimensionalen Raum zu bestimmen. Die abgetasteten Punkte werden gespeichert und die Ergebnisse werden über eine Software ausgegeben. Die 3D-Messarme werden nach der Anzahl der Drehachsen klassifiziert. In der Regel verfügen sie über 6 Drehachsen. Wenn sie zusätzlich einen Griff zur Steuerung des rotierenden Handgelenks haben, werden sie als Messarme mit 7 Achsen eingestuft [9]. Ein Beispiel für solche Messgeräte ist Faro Quantumm, das im Rahmen dieser Arbeit zur Erfassung der Messwerte eingesetzt wurde. Dabei erfolgte die Messung mit der 8-Achse. Die 8-Achse ist ein manuell drehbarer Drehtisch (siehe Abb. 1), der es ermöglicht, das zu messende Teil zu drehen und die Messung weiterhin im gleichen Koordinatensystem durchzuführen.



Abb. 1: Faro Quantum^{m} Arm (links) und 8-Achse Drehtisch (rechts) [10]

4. PRAKTISCHE DURCHFÜHRUNG

Einfluss von Stützstruktur-Parametern auf die Hirtisier®behandlung von LBM-Bauteilen

Für die Untersuchung wurden acht Proben (1-8) sowie für jede Probe ein Duplikat (9-16) mit unterschiedlichen Stützstruktur-Parameter-Kombinationen erstellt und geprüft. Hierfür wurde das Programm Magics verwendet. Die Duplikate sollten eine Beurteilung der Reproduzierbarkeit der Ergebnisse ermöglichen. Für die Auswertung der Ergebnisse werden für jeden Parametersatz (Probe und deren Duplikat) die arithmetischen Mittelwerte ermittelt. Um die Reproduzierbarkeit der Ergebnisse abzuschätzen, wird die Standardabweichung (STDEV) für jeden Parametersatz (PS 1-8) berechnet. Eine geringere Standardabweichung bedeutet, dass die Ergebnisse ein wenig von ihrem Mittelwert abweichen. Probe 15 (Duplikat der Probe 7) wurde fehlerhaft gedruckt und konnte daher nicht in den Versuch mit einbezogen werden. Der erste Parametersatz (PS1) dient in dieser Untersuchung als Referenz. Es wurden folgende Parameter variiert:

- Parametersatz 2 zielt darauf ab, die Auswirkungen eines geringeren Z-Offsets (siehe Abb. 2) zu untersuchen. Dafür wurden die Stützstrukturen im Parametersatz 2 mit einem kleineren Z-Offset als im Parametersatz 1 parametrisiert.

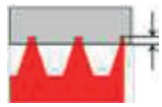


Abb. 2: Z-Offset [11]

- Die Proben 3 und 4 und ihre Duplikate wurden entworfen, um die Auswirkungen der eingefügten Fragmentierung auf die Festigkeit der Stützstruktur besser zu verstehen. Das X-Intervall sowie das Y-Intervall wurden für beide Proben gleich gewählt, während der Abstand zwischen den Fragmenten bei Probe 4 verringert wurde (siehe Abb. 3).

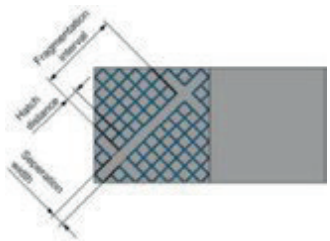


Abb. 3: Fragmentierungsparameter (Fragmentierungsintervall ="Fragmentation interval" und der Abstand (="Seperation width") und Hatch-Abstand [12]

- Die Perforationsparameter (siehe Abb. 4) wurden in die Proben 5 und 6 sowie deren Duplikate eingefügt, um herauszufinden, ob sie sich negativ auf die Stützstruktur-Festigkeit auswirken. Insbesondere sollte der Einfluss der Stegbreite untersucht werden. Zu diesem Zweck wurde die Stegbreite in Probe 6 reduziert.

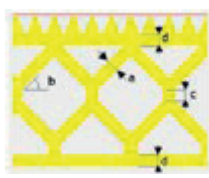


Abb. 4: Perforationsparameter: Stegbreite (a), Winkel (b), Höhe (c) und Solidhöhe (d) [13]

- Probe 7 unterscheidet sich von anderen Proben durch einen geringeren Hatch-Abstand, während die Fragmentierungs- und Perforationsparameter so gewählt wurden, dass sie denen des Parametersatzes 6 entsprechen, ebenso wie alle anderen Parameter. Diese Variante wurde explizit gewählt, um zu untersuchen, wie sich Hatch-Abstand (siehe Abb. 3) und Stegbreite (siehe Abb. 4) optimal kombinieren lassen.
- Der Parametersatz 8 soll ein besseres Verständnis der Auswirkungen von der Zahngeometrieparameter auf

die Festigkeit der Stützstruktur ermöglichen. Verglichen wurden Parametersatz 1 (ohne Zahnstruktur) und Parametersatz 8 (mit Zahnstruktur).

Einfluss von Innenstrukturen auf die Hirtisier®behandlung von LBM-Bauteilen

Für diese Untersuchung wurden drei Bauteilgeometrien mit annähernd gleichen Außenflächen, aber unterschiedlichen Verhältnissen der Innenfläche zur Gesamtoberfläche des Bauteils entworfen. Alle drei Probekörper haben demnach unterschiedliche Gesamtoberflächen. Es wurden mehrere Varianten dieser Probekörper gebaut. Eine Gruppe, bestehend aus je einer Kopie jeder Geometrie, sollte mit den gleichen Parametern behandelt werden. Anschließend sollte untersucht werden, ob der mittlere Abtrag an den jeweiligen Bauteilen auch mit dem Anteil der Innenfläche im Vergleich zur Gesamtoberfläche korreliert. Abb. 5 stellt das fertige 3D-Modell eines Probekörpers dar, das als Bauteiltyp 1 bezeichnet wird. Ziel bei der Konstruktion war es, ein Verhältnis von etwa 25% der Innenfläche im Vergleich zur Gesamtoberfläche zu erreichen. Das Bauteil wurde mit einem Anteil von 25,31% der Innenfläche im Vergleich zur Gesamtfläche entworfen.

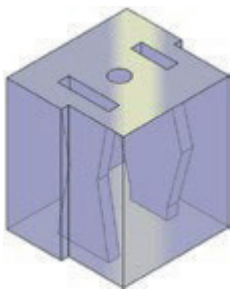


Abb. 5: Catia 3D-Modell: Bauteiltyp 1

Das Verhältnis der Innenfläche zur Gesamtfläche bei dem zweiten Bauteiltyp (siehe Abb. 6) beträgt 50,65%. Es wurden insgesamt fünf Kanäle konstruiert.

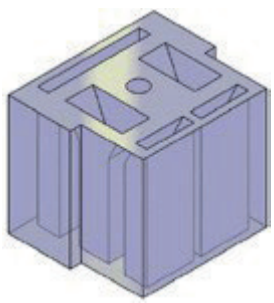


Abb. 6: Catia 3D-Modell: Bauteiltyp 2

Bei der Konstruktion des dritten Bauteiltypen (siehe Abb. 7) sollte im Idealfall 75% der Innenfläche im Vergleich zur Gesamtoberfläche erreicht werden. Hierbei haben die allgemein definierten Spezifikationen für die Hirtisier®behandlung einige Schwierigkeiten verursacht. Es wurden optimalen Abmessungen und Flächendaten ausgewählt, um ein Verhältnis von 65,75% der Innenfläche im Vergleich zu Gesamtfläche zu erreichen.

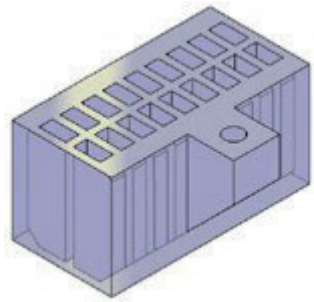


Abb. 7: Catia 3D-Modell: Bauteiltyp 3

Die Bauteile wurden mittels einer M6-Gewindestange kontaktiert, die in die Oberseite des Bauteils geschraubt wurde. Die Kontaktstellen von Gewindestange und Bauteil wurden bei allen drei Bauteilen durch Lackieren mit Enplate Stop-Off No. 1 (MacDermid Enthone) isoliert. Der Lack wurde vor der Behandlung über Nacht trocknen gelassen.
Abb. 8 zeigt die behandelten Bauteilen nach drei Hirtisier®-Prozessschritten. Es ist zu erwähnen, dass nach der letzten Behandlung ein zusätzlicher Schritt, das sogenannte „Dippen“, durchgeführt wurde, um eine glänzendere Oberfläche zu erhalten.

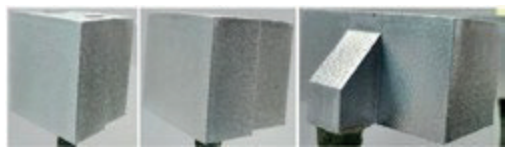


Abb. 8: Bauteil B (Typ 1, links), F (Typ 2, mittig) und J (Typ 3, rechts) nach dem Hirtisieren®

Entwicklung einer neuen Lösung für die Kontaktierung von Bauteilen

Es wurde versucht, die Bauteile im Elektrolyten so zu fixieren, dass sowohl Bauteile mit der für die Hirtisier®behandlung maximal spezifizierten Größe als auch kleinere Bauteile behandelt werden können. Die Idee, das Bauteil mit einem Draht aufzuhängen, wurde aufgrund des schnellen Verschleißes und der Notwendigkeit der regelmäßigen Entsorgung verworfen. Da das Bohren bei kleinen Bauteilen oft nicht praktikabel ist und auch nicht immer für die spätere Funktion des Bauteils vorgesehen ist, erweist sich das Klemmen von Bauteilen als die optimale Alternative. Dabei ist zu beachten, dass das Bauteil an den Klemmflächen nicht bearbeitet wird. Aus diesem Grund sollten die Klemmflächen möglichst klein sein. Die Klemmvorrichtung muss auch den Strom für die Behandlung zur Verfügung stellen. Die Stromzufuhr sollte an den Klemmstellen erfolgen, während der Rest der Klemme isoliert bleiben soll. Auf diese Weise werden zwei Vorteile erzielt. Zum einen wird die Klemme nicht aufgelöst und somit der Elektrolyt nicht verunreinigt, zum anderen kommen die Klemmstellen durch eine ausreichende Spannkraft nicht mit Elektrolyten in Berührung. Voraussetzung dafür ist, dass zwischen der Klemmfläche und dem Bauteil kein Zwischenraum vorhanden ist. Abb. 9 zeigt eine erste Entwurfsidee für die Klemme, die vier Druckfedern an den Stangen enthalten soll, um die Bauteile nach innen zu klemmen. Dieses Prinzip ermöglicht es, Bauteile unterschiedlicher Größe einzuspannen. Bei der

Konstruktion wurde darauf geachtet, dass alle Teile der Klemme mit geringem Aufwand hergestellt werden können.

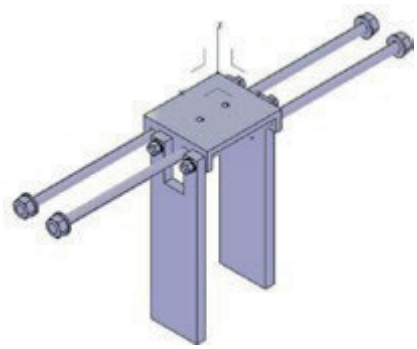


Abb. 9: Der erste Entwurf der Klemme – Isometrische Projektion

Um die Klemme möglichst universell einsetzbar zu machen, wurden die beiden vertikalen Platten asymmetrisch gestaltet, so dass sie sich gegenüberliegen und somit kongruent sind. Zusätzlich sind in der Hauptachse der vertikalen Platte mehrere versetzte Bohrlöcher angebracht, durch welche die Schrauben eingeschraubt werden (siehe Abb. 10). Je nach Anwendung und Bauteil wird eine andere Höhe verwendet. An dieser gewählten Stelle sollte eine längere Kontaktierungsschraube eingeschraubt werden, durch die der Kontakt eingeführt wird. Die Stelle, an der die Unterlegscheibe sitzt, muss dabei jedes Mal neu isoliert werden. Die Klemme sollte vollständig mit dem Isolierlack bedeckt sein, nur die Spitze der Kontaktierungsschraube sollte ausgenommen werden. Diese Spitze soll als Kontaktfläche dienen. Außerdem wurde ein entsprechendes Muster an den vertikalen Platten rechts und links von den Löchern konstruiert. Auf diese Weise wurde einerseits weniger Material verwendet, was zur Entlastung des Roboters führt, andererseits hat es den Vorteil, dass die Strömung des Elektrolyten nicht behindert wird, weshalb als Muster versetzte Dreiecke verwendet wurden. Für die Herstellung der vertikalen Platten werden Stahlbleche verwendet.

Die vier Stangen, an denen die Federn angebracht werden sollen, werden an beiden Enden mit Gewinden versehen und mit Muttern (M5 und M6) verschraubt. In Abb. 10 ist zu bemerken, dass die Stangen an beiden Enden einen Absatz haben. An der Stelle, an der die vertikale Platte befestigt ist, hat dieser Absatz die Funktion, vertikale Platte zu fixieren. Am anderen Ende der Stange befindet sich ein Gewinde, das zur Befestigung der Feder mit Hilfe einer Mutter dient. Hier wurde entsprechende Stufe mit dem Ziel konstruiert, um die Feder auf einer glatten Oberfläche ruhen zu lassen. Dadurch wird ausgeschlossen, dass die Feder an einem der Gewindeteile hängen bleiben könnte.

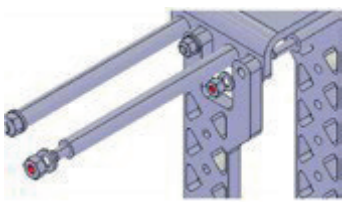


Abb. 10: Vier Stangen für die Druckfedern

Im unteren Teil der Klemme wurden noch zwei Platten angebracht. Sollte das Bauteil während der Behandlung verrutschen, sollen diese unteren Platten das Bauteil aufhalten und verhindern, dass es auf den Boden des Beckens fällt.

5. ERGEBNISSE

Parametersatz 2 weist im Vergleich zu Parametersatz 1 einen geringeren Z-Offset auf (siehe Tab. 1). Diese Parameteränderung wirkt sich positiv auf die Zugfestigkeit der Stützstrukturen aus, da laut den Ergebnissen für Parametersatz 2 eine höhere Zugkraft ausgeübt wurde.

Bauteil	Mittelwert der Zugkräfte (N)	Z-Offset (mm)	STDEV
PS1	1901,8	0,1	81,4
PS2	2562,3	0,03	110,5

Tab. 1: Ergebnisse des Zugversuchs: Z-Offset

Parametersatz 1, bei dem keine Fragmentierung gewählt wurde, und Parametersatz 3, bei dem die Fragmentierung mit einem Abstand von 0,8 mm erzeugt wurde, zeigen ähnliche Ergebnisse hinsichtlich der Zugfestigkeit. Bei Probe 4 wurde der Abstand auf 0,3 mm reduziert, was erwartungsgemäß zu einer höheren Zugfestigkeit führte. Alle Parameter, mit Ausnahme der Fragmentierungsparameter, wurden für diese drei Parametersätze gleich gewählt. Tab. 2 zeigt die gegenübergestellten Ergebnisse in Bezug auf die Fragmentierung.

	Mittelwert der Zugkräfte (N)	Fragmentierung - Abstand (c) (mm)	STDEV
PS1	1901,8	-	81,4
PS3	1905,9	0,8	116,3
PS4	2413,5	0,3	42,3

Tab. 2: Ergebnisse des Zugversuchs - Fragmentierung

Die Stützstrukturen im Parametersatz 3 sind überhaupt nicht perforiert, während im Parametersatz 5 und Parametersatz 6 eine Perforation mit unterschiedlicher Stegbreite (0,6 mm und 0,4 mm) gewählt wurde. Die Ergebnisse zeigen, dass das Einfügen der Perforation zu einer minimalen Absenkung der Stabilität führt. Das bedeutet, dass die Perforation eine positive Auswirkung auf die Hirtisier®barkeit hat. Wird jedoch die Stegbreite weiter reduziert, nimmt die Zugfestigkeit deutlich stärker ab. In Tab. 3 ist zu erkennen, dass die Zugfestigkeit des Parametersatzes 6 wesentlich geringer ist als die der beiden anderen betrachteten Parametersätze.

	Mittelwert der Zugkräfte (N)	Perforation - Stegbreite (a) (mm)	STDEV
PS3	1901,8	-	116,3
PS5	1793,5	0,6	46,7
PS6	824,7	0,4	53,1

Tab. 3: Ergebnisse des Zugversuchs - Perforation

Die Tab. 4 zeigt eine höhere Zugfestigkeit für Probe 7, was darauf hindeutet, dass ein geringerer Hatch-Abstand zu einer deutlich höheren Zugfestigkeit führt. Es ist zu beachten, dass die durch Stegbreite reduzierte Zugkraft im Parametersatz 6 durch einen geringeren Hatch-Abstand kompensiert werden kann. Durch diese Kombination kann der Elektrolyt leichter durch die Perforation in die Stützstruktur eindringen und somit diese leichter entfernen, während die notwendige Festigkeit

der Stützstruktur durch einen geringeren Hatch-Abstand erhalten bleibt.

	Mittelwert der Zugkräfte (N)	Hatch-Abstand (mm)	STDEV
PS6	824,7	1	53,1
PS7	1683,0	0,6	-

Tab. 4: Ergebnisse des Zugversuchs - Hatch-Abstand

Die Probe 1 weist eine höhere Zugfestigkeit als die Probe 8 auf. Es ist zu beachten, dass der Parametersatz 8 eine höhe Standardabweichung im Vergleich zu anderen Parametersätzen hat. Die hinzugefügte Zahnstruktur führt daher nicht unbedingt zu einer niedrigeren Stabilität.

	Mittelwert der Zugkräfte (N)	Zähne oben - Höhe (a) (mm)	STDEV
PS1	1901,8	1	81,4
PS8	1645,4	0,6	177,6

Tab. 5: Ergebnisse des Zugversuchs - Zahngeometrie

Tab. 6 zeigt die zusammengefassten Ergebnisse für alle drei Bauteile. Es wurde der mittlere Abtrag von 217-238 µm ermittelt. Die beiden gemessenen Werte je Bauteil unterschieden sich nicht stark voneinander (1-4 µm), aber von Bauteil zu Bauteil war die Abweichung deutlich höher. Die Ergebnisse zeigen, dass der mittlere Abtrag indirekt mit dem Anteil der Innenfläche an der Gesamtfläche korreliert (je größer der Anteil, desto geringer der mittlere Abtrag an der Außenfläche). Somit stimmen die Ergebnisse mit den bisherigen Erkenntnissen aus den vorangegangenen Versuchen überein. Die maximale Abweichung des mittleren Abtrags von einem Bauteil zum anderen beträgt in diesem Versuch 25 µm. Der Unterschied im mittleren Abtrag vom Bauteil B (Anteil Innenfläche an Gesamtfläche: 25%) im Vergleich zu dem Bauteil F (Anteil Innenfläche an Gesamtfläche: 51%) und dem Bauteil J (Anteil Innenfläche an Gesamtfläche: 66%) kann als signifikant betrachtet werden (17-21 µm). Die Abweichung im mittleren Abtrag von Proben F und J (4 µm) liegt allerdings im Bereich der Abweichung der beiden je Bauteil gemessenen Abstände und muss daher als nicht signifikant betrachtet werden. Ähnliches wurde bei einer erneuten Auswertung des vorangegangenen Versuchs festgestellt.

Bauteil	Mittlerer Abtrag (mm)	Anteil der Innenfläche an Gesamtoberfläche (%)
1	0,238	25
2	0,221	51
3	0,217	66

Tab. 6: Mittlerer Abtrag und Anteil der Innenfläche an Gesamtfläche für alle drei Bauteilen

Eine visuelle Darstellung der Korrelation zwischen dem mittleren Abtrag und dem Anteil der Innenfläche im Vergleich zur Gesamtfläche ist in Abb. 11 zu sehen.

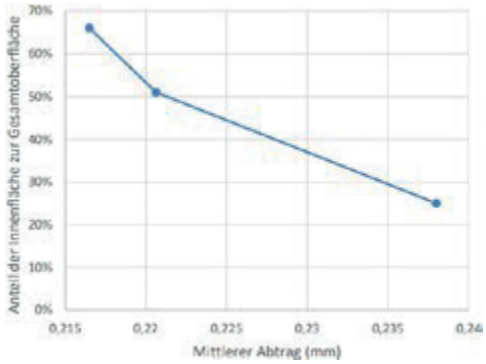


Abb. 11: Mittlerer Abtrag in Abhängigkeit von dem Anteil der Innenfläche im Vergleich zu Gesamtfläche

Abb. 12 zeigt eine endgültige Konstruktionsidee für die Klemme, die aus fünf einzelnen Blechen sowie vier Stäben besteht.

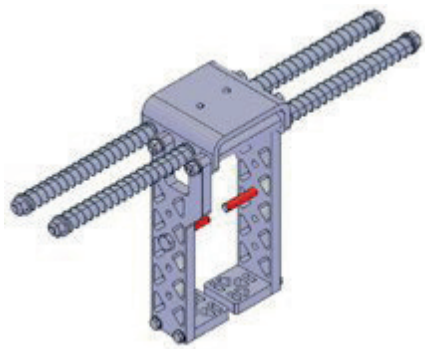


Abb. 11: Fertige Klemme

Die Klemme wurde mittels Laserschneiden aus rostfreiem Blech (Werkstoff 1.4301) gefertigt werden. Als Material wurde rostfreier Stahl und nicht Aluminium gewählt, aufgrund der möglichen Verbiegung des Aluminiums. Außerdem sorgt der Stahl für eine längere Lebensdauer der Gewinde. Aus fertigungstechnischen Gründen wurden die abzuschneidenden scharfkantigen Stellen abgerundet. Es ist darauf zu achten, dass die Kontakte durch die Federn und Stäbe in den Bohrlöchern nicht ausreichen und dass es bei hohen Strömen sogar zum Verschweißen kommen kann. Deshalb sollten die beiden vertikalen Platten jeweils mit einem Kabel mit dem Warenträger oder zumindest mit der Halterung verbunden werden. Zusätzlich werden auch Gleitlager aus Kunststoff in die Bohrungen der Halterung eingesetzt werden.

6. ZUSAMMENFASSUNG UND AUSBLICK

In dieser Masterarbeit wurden die Ergebnisse der hirtisiert®en Bauteile untersucht und analysiert, um die oben gestellten Forschungsfragen zu beantworten. Das Hirtisier®-Verfahren hat sich als sehr gut für die Nachbehandlung von

LBM-Bauteilen erwiesen. Die Oberflächen der Bauteile wurden geglättet, und die Stützstrukturen wurden entfernt. Darüber hinaus bietet diese neu entwickelte Nachbehandlungsmethode Verbesserungsmöglichkeiten und sollte für verschiedene Anwendungen weiter angepasst werden. Bei der Untersuchung der Festigkeitseigenschaften von Zugproben wurde eine interessante Erkenntnis gewonnen: Die geringere Zugfestigkeit aufgrund einer kleineren Stegbreite kann durch einen geringeren Hatch-Abstand kompensiert werden. Diese Kombination ist für das Hirtisieren® gut geeignet, da eine geringere Stegbreite die Entfernung der Stützstrukturen im Elektrolyten erleichtert. Durch den verringerten Hatch-Abstand wird sichergestellt, dass die Stützstrukturen ausreichend stabil bleiben. Diese Kombination ist jedoch keine optimale Lösung für die Geometrie der Stützstrukturen. Vielmehr sollte ein größerer Hatch-Abstand mit einer höheren Stegbreite sinnvoller sein. In den künftigen Versuchen soll daher untersucht werden, ob die Vorteile der inversen Kombination beibehalten werden können.

Bei der Untersuchung der Auswirkungen unterschiedlicher Innenflächen im Verhältnis zur Gesamtoberfläche auf das Hirtisieren® wurde festgestellt, dass die geforderten Bauteilspezifikationen die Designfreiheit des 3D-Drucks einschränken. Dennoch wiesen die Bauteile einen ausreichenden prozentualen Unterschied im Verhältnis der Innenfläche zur Gesamtoberfläche auf, und die Untersuchung konnte erfolgreich durchgeführt werden. Es wurde festgestellt, dass der Abtrag nicht vollständig homogen über das einzelne Bauteil verteilt ist. Zwar wurden von jedem Bauteil die Stützstrukturen entfernt und die Innenflächen geglättet, jedoch wurden die Außenflächen stärker abgetragen. Darüber hinaus wurde beobachtet, dass der Unterschied im mittleren Abtrag nur dann ausgeprägt ist, wenn der Anteil der Innenfläche im Vergleich zur Gesamtoberfläche unter 40% liegt. Diese Annahme sollte in einem weiteren Versuch geprüft werden.

Die in dieser Arbeit entwickelte Klemmvorrichtung wurde im Rahmen des elchemPP-Projekts weiter gefertigt und getestet. Obwohl sie derzeit als schnelle und einfache Kontaktierung von kleineren Bauteilen eine optimale Lösung für künftige Untersuchungen darstellt, ist eine Weiterentwicklung dieser Klemme nicht ausgeschlossen.

7. LITERATURVERZEICHNIS

[1] W. Hansal, S. Hansal, R. Mann und G. Sandulache, "Electropolishing method and system therefor,"

[2] F. Brenne, *Selektives Laserschmelzen metallischer Materialien: Einfluss von Prozessparametern und Miniaturisierung auf Mikrostruktur und mechanisches Verhalten geometrisch komplexer Strukturen*. Kassel: Kassel University Press, 2018.

[3] K. Zeng, *Optimization of support structures for selective laser melting*, 2015.

[4] RENA Technologies GmbH, *Automatisches Post-Processing 3D-Gedruckter Metallbauteile.* [Online]. Verfügbar unter: https://www.rena.com/de/produkte/additive-manufacturing/finishing-module/.

[5] R. Triebnik, Hg., *H 3000 FINISHING MODUL: Betriebsanleitung für Anlage*, 2017.

[6] FOTEC Forschungs- und Technologietransfer GmbH, „HIRTISATION OF ALM PARTS: Design & Treatment Guidelines". Unveröffentiches Dokument, 2021. Zugriff am: 3. Mai 2021.

[7] EOS GmbH - Electro Optical Systems, „EOS Aluminium AlSi10Mg: Materialdatenblatt", 2015.

[8] M. Launhardt *et al.*, „Detecting surface roughness on SLS parts with various measuring techniques", *Polymer Testing*, Jg. 53, S. 217–226, 2016, doi: 10.1016/j.polymertesting.2016.05.022.

[9] FARO Europe GmbH & Co. KG, *3D-Messarmtechnologie: Technologie-Whitepaper.* [Online]. Verfügbar unter: https://media.faro.com/-/media/Project/FARO/FARO/FARO/Resources/2021/01/29/00/44/Whitepaper-Der-menschliche-Arm-als-Vorbild-in-der-Koordinatenmesstechnik-DEU.pdf?rev=780b637e78af4845b90ae7e4c6f67baa.

[10] FARO, „FARO® Quantum Max ScanArms", 2023. [Online]. Verfügbar unter: https://www.faro.com/en/Products/Hardware/ScanArms.

[11] P. N. Johannes Lindecke, H. Blunk, J.-P. Wenzl, M. Möller und C. Emmelmann, „Optimization of support structures for the laser additive manufacturing of TiAl6V4 parts", *Procedia CIRP*, Jg. 74, S. 53–58, 2018, doi: 10.1016/j.procir.2018.08.029.

[12] W. Ameen, M. Khan Mohammed und A. Al-Ahmari, „Evaluation of Support Structure Removability for Additively Manufactured Ti6Al4V Overhangs via Electron Beam Melting", *Metals*, Jg. 9, Nr. 11, S. 1211, 2019, doi: 10.3390/met9111211.

[13] M. Pasterk, „Entwicklung einer CAx-Prozesskette für die additive Fertigung von Bauteilen aus Metall", Technischen Universität Wien: Fakultät für Maschinenwesen und Betriebswissenschaften, 2018.



Ana Kokic, BSc
Pottendorfer Straße 62, 2700 Wiener Neustadt, Österreich
06601599892, anakokic30@gmail.com

Geburtsdatum: 26.06.1998

Bildungsweg:

2020 –2023 Fachhochschule Wiener Neustadt, Masterstudium Mechatronik

2017 –2020 Fachhochschule Wiener Neustadt, Bachelorstudium Wirtschaftsingenieurwesen

Praktische Erfahrung:

Praktikantin im Bereich Operations & Service, EBG MedAustron GmbH

Studienassistentin im Fachbereich Electrical Engineering, Fachhochschule Wiener Neustadt

Praktikantin im Bereich Environment, Health & Safety, Ardagh Metal Beverage Manufacturing Austria GmbH

Mitarbeiterin im Verkaufsinnendienst, Herz Armaturen GmbH

Praktikantin im Bereich Vertragsmanagement, Trans Austria Gasleitung GmbH

Sprachen: Serbokroatisch (Muttersprache), Deutsch (C1), Englisch (C1)

Hobbies: Lesen, Reisen, Sport, Moderne Physik, Klavier spielen

ENTWICKLUNG EINES PROTOTYPEN ZUM AUTOMATISIERTEN VERZWIRNEN VON WOLLE FÜR DEN PRIVATEN GEBRAUCH

Paul Isaak Leitner BSc.

Kurzfassung: *Das Ziel der Arbeit ist es, ein Gerät für den privaten Gebrauch zu entwickeln, welches in der Lage ist, zwei bis vier handgesponnene Wollfäden automatisiert zu verzwirnen. Die Anforderungen dazu Umfassen die Regelung der Motoren für einen gleichmäßigen Zwirn, das gleichmäßige Aufwickeln des erstellten Zwirns auf einer bereitgestellten Spule und Erfassung einiger Daten während des Zwirnens wie die Fadenlänge und den Zug, der auf den Fäden lastet. Die Maschine wird mithilfe von mehreren Motoren einen Mikro-Controller und verschiedenen Sensoren betrieben. Der am Ende der Arbeit getestete Prototyp war in der Lage, unter einigen Auflagen zu der Qualität der Fäden, einen Zwirn zu produzieren. Dabei wurden auch fast alle Anforderungen erfüllt und für nicht erfüllte Aufgaben wie die Lautstärke oder der Geschwindigkeit des Zwirnens wurden Lösungen vorgeschlagen.*

Schlüsselwörter: *Zwirnmaschine, Maschinen-Konstruktion, einfache Motorregelung*

1. EINLEITUNG

In letzter Zeit beschäftigen sich immer mehr HandarbeiterInnen mit den erstellen ihrer „Rohware“ also den Spinnen und Verzwirnen. Normalerweise werden fertige Knäuel Wolle oder Zwirn zum Häkeln, Stricken oder anderen Handarbeitstechniken verwendet. Beim Spinnen geht man noch einen Schritt weiter zurück und erarbeitet sich dieses Knäuel Wolle selbst. Doch während das Spinnen selbst ein spannender und konstruktiver Prozess ist, ist das Verzwirnen mit der Hand oder mit dem Spinnrad jedoch eher eintönig. Aus diesem Grund kommt es öfters vor, dass eine HandarbeiterIn ihren Wunsch, eine eigene Schnur für ihre Strickwahren zu verwenden bei dem Prozess des Verzwirnens wieder aufgibt. Im Rahmen dieser Arbeit wird daher ein Gerät entwickelt, welches den Prozess des Zwirnens automatisiert. Natürlich ist das Spinnen und Verzwirnen längst maschinell gelöst - diese Maschinen sind aber in der Regel Industriemaschinen die wegen Größe, Lärm und schwerer Bedienbarkeit nicht für eine Anwendung im heimischen Wohnzimmer in Frage kommen. Der Fokus bei der Entwicklung dieser Arbeit ist daher, den Prozess des Zwirnens möglichst einfach und kostengünstig zu realisieren. Im Gegenzug dazu wird das Gerät nicht auf hohe Wirtschaftlichkeit und Produktionsgeschwindigkeit optimiert werden. Das Zwirngerät soll eine Möglichkeit haben zwei oder mehr handelsübliche Spulen eines Spinnrades mit einem handgesponnenen Faden aufzunehmen, deren Fäden ineinander zu verdrehen und auf einer weiteren Spule wieder aufzurollen. Dabei wird die Qualität des Fadens überwacht und durch richtige Regelung der Motoren sichergestellt.

2. AUFGABENSTELLUNG

Die Aufgabenstellung dieser Arbeit ist es, zwei bis vier Wollfäden mit einer zufriedenstellenden Qualität zu verzwirnen und nach dem Vorgang ordentlich auf einer weiteren Spule aufzuwickeln. Zufriedenstellende Qualität bedeutend erstens einen bestimmten Drall in jeder Stelle des fertigen Zwirns. Zweitens sollen sich die Fäden gleichmäßig umeinanderschlingen. Eine weiter Aufgabe ist das Erkennen eines Fadenriss oder des Erreichens des Ende eines Fadens, sowie das Anhalten des Geräts nach Detektion eines solchen Falls.

Neben den Aufgabenstellungen in Bezug auf die Qualität des Faden selbst ergeben sich auch einige Anforderungen an das Gerät. Da es für den nicht kommerziellen Endnutzer gedacht ist muss es möglichst einfach zu bedienen sein, auf einen Tisch passen und während des Betriebs nicht zu laut werden.

3. MATERIALIEN UND METHODEN

Beim Vorgang des Spinnens werden lose Fasern durch Verdrehen miteinander verbunden und so eine endlose längliche Struktur (Garn) erschaffen. Durch das Verdrehen bekommen die einzelnen Fasern genug Halt zueinander, dass sie sich nur noch schwer voneinander lösen können. Der Begriff Garn ist dabei in DIN 60900 definiert als Sammelbegriff für alle linienförmigen textilen Gebilde. Von einem Faden darf man dabei sprechen, wenn man ein begrenztes Stück eines Garnes meint. [Claßen 2009]

Ein einzelner Faden für sich verliert allerdings seine Stabilität, sollte er sich wieder aufdrehen. Eine Möglichkeit, um ein Aufdrehen des Garns zu verhindern ist das Zwirnen. Bei diesem Vorgang werden mindestens zwei Einzelgarne noch einmal miteinander verdreht. Allerdings in die andere Richtung, als zuvor gesponnen wurde. Diesen Vorgang würden zwei lose Fäden „von selbst“ machen, wenn sie nebeneinandergehalten werden, da sie sich, wenn sie sich entspannen und aufdrehen, ineinander verdrehen. Es sollte beim Verzwirnen auch darauf geachtet werden, nur so viel Drehung zu geben, wie auch der Garn beim Spinnen erhalten hat [Claßen 2009]. Durch das Zwirnen wird der Drall in den Fäden gespeichert. Selbst wenn er sich wieder aufdrehen sollte, liegen immer noch die zwei einzelnen Fäden vor, die sich wieder eindrehen, sobald sie nicht mehr daran gehindert werden.

Für eine Spinn- oder Zwirnmaschine ist wichtig wieviel Drall ein Garn benötigt. Sehr feine Fäden brauchen dabei mehr Drall, dickere weniger. Auch das spätere Einsatzgebiet hat Einfluss auf den nötigen Drall [Claßen 2009, Vogt 2008]. Eine Variable Einstellung des Dralls ist für die Zwirnmaschine also notwendig, um verschiedene Wolldicken verarbeiten zu können.

4. PRAKTISCHE DURCHFÜHRUNG

Der Zwirner wurde im ersten Schritt mittels CAD-Software (Inventor) geplant und danach aus Buchenholzplatten sowie einigen 3D-Druck Bauteilen zusammengebaut. Die einzelnen Baugruppen werden mithilfe von zwei Stepper-Motoren und ein DC-Motor angetrieben. Für die Überprüfung der Wollqualität ist eine Wägezelle für die Zugmessung und ein Encoder für die Längenmessung des Zwirns in Verwendung. Die Steuerung und Regelung übernimmt der Micro-Controller ESP32. welcher auf einer Lochrasterplatte gemeinsam mit den Motortreibern für die Stepper-Motoren verlötet wurde.

4.1 Aufbau

Der Zwirner besteht aus drei mechanischen Hauptbaugruppen, die auf einem Chassis aus Holz gelagert werden. Diese sind, wie auf Abbildung 1 oben zu erkennen, die Zwirnvorrichtung, die Aufwickeleinheit und die Fadenführung. Alle drei Baugruppen werden von je einem Motor angetrieben, der hinter der Rückwand befestigt und gesteuert wird, wie auf Abbildung 1 unten zu sehen ist.

Die Zwirnvorrichtung dient dabei zur Aufnahme der Spulen mit den gesponnenen Fäden. Die Fäden werden aus der Zwirnvorrichtung hinausgeführt. Durch die Drehung der Zwirnvorrichtung wird Drall in die Fäden gebracht und so verzwirnt.

Der Zwirn wird nun über zwei Bauteile umgelenkt, welche den Zug des Fadens und die Länge des Fadens messen. Der Zug des Fadens wird über eine Wägezelle [Tränkler 1992] gemessen. Damit kann ermittelt werden, ob noch alle Fäden aus der Zwirnvorrichtung gezogen werden oder ob einer gerissen ist. Die Länge des Fadens wird über eine Rolle gemessen, um die der Zwirn einmal herumgewickelt wird. So bewegt sich die Rolle mit dem Zwirn mit und über einen Encoder kann Geschwindigkeit und Länge ermittelt werden (Meterzähler).

Das Aufwickeln des fertigen Zwirns und daher auch das Herausziehen aus der Zwirneinheit, übernimmt die Aufwickeleinheit. Wichtig dabei ist, dass die Geschwindigkeit des Fadens immer die gleiche bleibt, damit auch der Drall im Faden gleichmäßig ist. Die Fadenführung sorgt dafür, dass der Zwirn gleichmäßig auf der Aufwickelspule aufgewickelt wird, indem sie ihn auf einen Umlenkhaken umlenkt, der sich entlang der Spule bewegt.

Im Antriebsteil (Abbildung 1 unten) erkennt man einerseits die drei Motoren, wobei zwei davon baugleiche Stepper Motoren sind, und der andere ein DC-Motor. Der DC-Motor treibt die Zwirnvorrichtung an, die beiden anderen steuern den Aufwickler und die Fadenführung. Zu sehen ist außerdem eine Lochraster-Platine mit dem Mikrocontroller ESP32, welcher die gesamte Steuerung der Maschine übernimmt. Auf der Platine sind noch zwei Motortreiber (DRV8825) angebracht, welche die Stepper Motoren ansteuern. Außerdem beherbergt sie die Ausgänge für sämtliche Sensoren und Schalter und die Anschlüsse für das Display und den Bedien-Button, über den das Gerät bedient und gestartet werden kann.

4.2 Regelung und Ablauf des Zwirnes

Um im Programm die Motoren richtig ansprechen zu können, um so einen Zwirn in guter Qualität herzustellen, werden zuerst der Soll-Drall D benötigt, mit dem der die Fäden gezwirnt werden soll. Mit diesen Parametern und den laufenden Werten der Sensoren und Encoder kann die nötige Geschwindigkeit der Motoren berechnet und geregelt werden.

Zu Beginn eines Zwirnvorgangs wird der DC-Motor auf eine vordefinierte Soll-Geschwindigkeit n_{Zsoll} gebracht. Mithilfe der Ist-Geschwindigkeit n_{Zist}, die über einen eingebauten Inkrementalencoder [Schmusch 1993] mitgemessen wird, den gewünschten Drall und den Umfang der leeren Spule U_{Spule} kann die Soll-Geschwindigkeit für den Aufwickelmotor n_A errechnet (1) und ausgegeben werden.

$$n_A = n_{Zist}/D * U_{Spule} \quad (1)$$

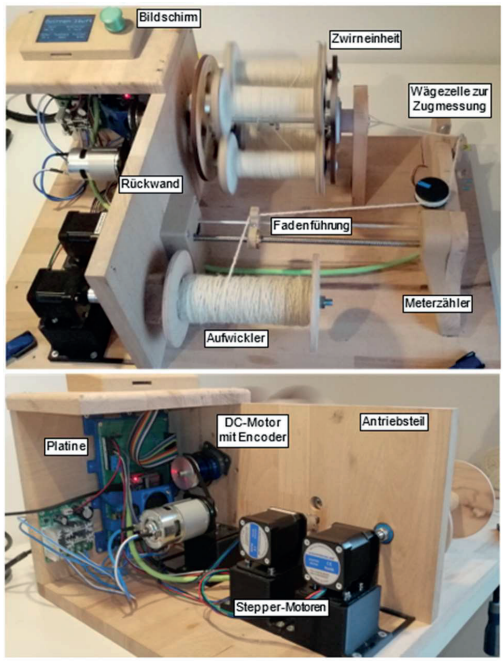


Abb. 1: Aufbau des Zwirners

Diese Regelung funktioniert allerdings nur, solange die Spule noch leer ist, denn sobald die zweite Schicht Wolle auf die Spule gewickelt wird, ändert sich der Umfang der Spule. Der Aufwickler würde somit immer schneller und so immer weniger Drall in den Faden gebracht werden.

Deshalb wird, sobald der Zwirner seine Soll-Geschwindigkeit erreicht hat, die Aufwickelgeschwindigkeit mithilfe des Meterzählers errechnet. Im Programm wird gemessen, wie lange der Encoder einen für einen Schritt benötigt t_{ist}, was 5mm Zwirn entspricht. Ebenfalls mit der Ist-Geschwindigkeit des Zwirners und dem Soll-Drall kann die Soll-Zeit t_{soll} errechnet werden, wie lange für 5mm Zwirn gebraucht werden soll. Durch Vergleichen von t_{ist} und t_{soll} kann die Geschwindigkeit des Aufwickelmotors gesteuert werden.

$$t_{soll} = n_{Zist}/D * 2 \quad (2)$$

5. ERGEBNISSE

Der Prototyp konnte beweisen, dass ein maschinelles Zwirnen mit einem kleinen, für den Privathaushalt geeigneten Gerät möglich ist. So ist es am Ende der Arbeit möglich zwei bis vier handgesponnene Fäden zu verarbeiten. Beispiel dafür sind auf Abbildung 2 zu sehen. Die einzige Einschränkung, die während der Versuche aufgefallen ist, ist die Belastbarkeit der Einzelfäden. Da das Ausziehen der Fäden aus der Zwirneinheit im laufenden Betrieb eine gewisse Kraft benötigt, müssen die Fäden diesen Kräften auch standhalten können. Die übliche Stärke von handgesponnen Fäden erfüllt für gewöhnlich diese Anforderung. Lediglich sehr dünne Fäden können mit dieser ersten Version des Zwirners nicht verarbeitet werden.

Ebenso wurden die Ziele zur Qualität und Genauigkeit bei der Drall-Regelung erreicht. Die verzwirnten Wollknäuel sind nicht von händisch gezwirnten Produkten zu unterscheiden. Zu Gunsten der Qualität wurden jedoch die Ziele zur Geschwindigkeit knapp verfehlt. Die Zwirndauer für 100m Zwirn lag immer leicht über den vorgegebenen 30 min.

Die weiteren Ziele zu Größe und Bedienbarkeit wurden in ihrer Grundanforderung erfüllt. Die Grundplatte der Maschine hält sich an die Größen-Anforderung und auch die maximale Höhe wurde nicht überschritten. Die Verwendung des Geräts ist sehr einfach gestaltet und sie ist mit Spulen einspannen, Faden einfädeln und Einstellungen treffen in wenigen Minuten einsatzfähig.

Bei diesen letzten Punkten zeigt sich jedoch sehr stark, dass es sich bei der Maschine um einen Prototyp handelt. So wurde die Größenanforderung zwar eingehalten, jedoch ist sehr gut ersichtlich, dass hier noch viel Platz eingespart werden kann und weitere Versionen deutlich kleiner designt werden könnten. Auch kann die Bedienung und das Vorbereiten der Maschine zum Zwirnen deutlich einfacher gelöst werden. Beispielsweise müssen bei der Zwirneinheit vor jedem Einsatz mehrere Muttern angezogen und Kugellager aus- und eingebaut werden, um die Spulen einsetzen zu können.

Das letzte zu erreichende Ziel ist die Lautstärke des Geräts während des Betriebs. Hier wurde die Bewertung in zwei Teile unterteilt. Der Betrieb des Zwirners selbst (bei Nichtbeachtung der Motoren) ist in einer angemessenen Lautstärke. Hier konnten während der Entwicklung auch klare Fortschritte gemacht werden, denn bei den ersten Versuchen verursachte vor allem die Zwirneinheit ein störendes, lautes Klopfen.

Die Stepper Motoren sind im laufenden Betrieb aber sehr laut und könnten in dieser Form nicht in einem fertigen Produkt verwendet werden. Somit benötigt dieser Teil der Maschine noch eine klare Verbesserung und dieses Ziel wurde am Prototyp nicht erreicht.



Abb. 2: Mit den Zwirner produzierte Wolle

6. ZUSAMMENFASSUNG UND AUSBLICK

Während der Entwicklung des Zwirners wurden viele Erkenntnisse gewonnen, wie das Gerät in einer nächsten Version weiter verbessert werden könnte.

Die Fadenführung erschwert durch ihre derzeitige Position die genaue Regelung des Dralls. Bei der Bewegung Richtung Rückwand wird zusätzlicher Faden ausgezogen und bei der Fahrt in die andere Richtung wird dieser zusätzliche Faden wieder zurückgegeben. Somit verändert sich die Auszugsgeschwindigkeit immer bei einem Richtungswechsel der Fadenführung. Wenn die Anordnung der gesamten Aufwickeleinheit sich um 90° dreht, wird dieser Effekt sehr stark minimiert.

Der verwendete Stepper Motor der Aufwickeleinheit war bei zunehmendem Spulenfüllstand auch zu schwach, um den Faden weiter aus der Zwirneinheit zu ziehen und aufzuwickeln. Hier wurde der steigende Kraftbedarf durch den größer werdenden Hebel unterschätzt. Die Kraft des Motors reicht aus, um 100g Wolle zu verarbeiten. Einer der Vorteile des Zwirners wäre aber die Fähigkeit, größere Knäuel zu produzieren. Dazu muss bei einer weiteren Version des Zwirners der Motor leistungsstärker dimensioniert werden. Hier könnten Versuche mit einem DC-Motor gemacht werden.

Von Vorteil wäre auch, wenn der Motor der Zwirneinheit sich in beide Richtungen drehen könnte. Der bisherige Treiber kann den Motor nur im Uhrzeigersinn bewegen. Für das Zwirnen von Wolle ist das in den meisten Fällen passend. In einigen Situationen muss jedoch gegen den Uhrzeigersinn gedreht werden, weswegen ein ausgereiftes Produkt das auch anbieten sollte.

Die letzte Erkenntnis, betrifft die Möglichkeiten, die Belastung der Fäden auf ihrem Weg durch die Maschine zu verringern. Zum einen kann eine Entlastung durch glattere und optimierte Umlenkrollen und größere Durchführungen erreicht werden, wodurch weniger Reibung auf den Faden wirkt. Eine weitere Möglichkeit ist ein geschickteres Abwickeln der Spulen in der Zwirneinheit. Hier könnte der maximale Winkel in der der Faden von der Spule abgewickelt wird, durch eine Drehung der Spulen um 90° erreicht werden.

7. LITERATURVERZEICHNIS

Claßen-Büttner U., 2009, *Spinnst Du? Na klar!,* Book on Demand GmbH, 978-3-8391-1752-2, Norderstedt

Vogt S., 2008, *Geschichte und Bedeutung des Spinnrads in Europa,* Shaker Media, 987-3-86858-074-7, Aachen

Tränkler H.-R., 1992, *Taschenbuch der Meßtechnik,* R. Oldenburg Verlag GmbH, 3-486-22230-9, München

Schmusch W., 1993, *Elektronische Meßtechnik Prinzipien, Verfahren, Schaltungen,* Vogel Verlag und Druck KG, 3-8023-1494-8, Würzburg



Paul Isaak Leitner BSc.

Hauptstraße 38/1, 2564 Weissenbach an der Triesting; Tel. 0660/513 1004; E-Mail: paul.isaak.leitner@gmx.at

Ausbildung zum Gärtner. Nach einigen Jahren Ausübung dieses Berufs, Studium an der FH Wiener Neustadt, Fachrichtung Mechatronik. 2020 Bachelorabschluss, danach Masterstudium der derselben Fachrichtung. Seit vielen Jahren beschäftige ich mich zudem hobbymäßig mit Spinnen und allem, was dazu gehört, wodurch ich auf die Idee mit den Zwirner gekommen bin.

ENTWICKLUNG INTELLIGENTER ERDANKERSENSORIK ZUR MESSUNG DES FEUCHTEGEHALTES UND DER FARBE VON BETON

Mustafa Oktay

Kurzfassung: *Die Überwachung von Bauteilen, die eingegraben oder eingebettet sind, kann schwierig sein, da sie schwer zugänglich sind. Es gibt Verfahren, die für die Überprüfung solcher Bauteile genutzt werden, die oft aufwendig und teuer sind. In dieser Arbeit wird eine Smart-Sensorik entwickelt, um diese Herausforderungen zu lösen. Dazu wurde ein Feuchte- und Farbsensor entwickelt und getestet. Für die Entwicklung der Smart-Sensorik wurden theoretische Grundlagen und Randbedingungen berücksichtigt. Eine anwendungsnahe Simulation wurde erstellt, um die Umsetzbarkeit und Funktionsweise des Sensors zu überprüfen. Anschließend wurde ein Prototyp hergestellt und unter Laborbedingungen getestet. Die Messdaten werden von dem Sensor auf eine Datenbank hochgeladen und können jederzeit über eine Webapplikation abgerufen werden. Die Webapplikation ermöglicht eine Steuerung des Sensors sowie die Auswertung und Speicherung der Messdaten als Graphen beziehungsweise Tabellen.*
Schlüsselwörter: *Eingebettet, Smart-Sensorik, Simulation, Webapplikation, Bauindustrie*

1. EINLEITUNG

Diese Arbeit beschäftigt sich mit der Entwicklung eines Feuchte- und Farbsensors, der in der Bauindustrie zur ferngesteuerten Überwachung von eingegrabenen oder eingebetteten Bauteilen eingesetzt werden kann. Die Arbeit nutzt grundlegende Konzepte des IoT (Internet of Things) und der Industrie 4.0, um industrielle Prozesse mit digitaler Informations- und Kommunikationstechnologie zu verknüpfen. Die Arbeit steht im Kontext der digitalen Transformation und zeigt, wie die Anwendung von Sensortechnologie und vernetzten Systemen auch in der Bauindustrie möglich ist. Das Ziel dieser Arbeit ist es, einen Sensor zu entwickeln, der unter schwierigen Umgebungsbedingungen, wie zum Beispiel Einflüsse durch Staub und Schmutz auf Baustellen zuverlässig arbeitet und feuchte- und farbabhängige Daten erfasst. Die Sensoren werden in einem Gehäuse platziert, das in der Nähe des Messobjekts zu liegen kommen, ohne direkten Materialkontakt herzustellen.

Die Messdaten sollen dann an eine zentrale Stelle gesendet werden, um eine effizientere Überwachung und Kontrolle der Bauprozesse zu ermöglichen.

Um die Ziele dieser Arbeit zu erreichen, werden Literaturrecherchen, vergleichende Studien und Experimente durchgeführt. Die Ergebnisse dieser Arbeit können helfen, die Effizienz und Sicherheit im Bauprozess zu verbessern, indem eine kontinuierliche Überwachung und Steuerung ermöglicht wird.

Insgesamt zeigt die Arbeit den innovativen Charakter des Einsatzes von IoT-Technologien in der Bauindustrie und unterstreicht die Bedeutung von vernetzten Sensoren für die Überwachung und Steuerung von Prozessen in der Zukunft.

2. PROBLEM- UND AUFGABENSTELLUNG

Ein Problem in der Baubranche ist die Notwendigkeit, die Feuchtigkeit und Farbe von Bauteilen zu überwachen. Feuchtigkeit und Verfärbungen können auf Schäden oder Materialermüdung hinweisen, die zu einem Bauteilversagen führen können. Daher besteht die Herausforderung, Sensoren zu entwickeln, die in der Lage sind, die Feuchtigkeit und Farbe von Bauteilen automatisch und zuverlässig zu überwachen. Solche Sensoren können helfen, potenzielle Schäden frühzeitig zu erkennen und zu vermeiden, indem sie Echtzeitdaten an ein zentrales System senden. Die Herausforderung besteht darin, Sensoren zu entwickeln, die robust genug sind, um schwierigen Umgebungsbedingungen standzuhalten, aber gleichzeitig kostengünstig und einfach zu installieren sind.

3. MATERIALIEN UND METHODEN

Die theoretischen Grundlagen für Feuchtigkeitsmessung sind ein wichtiger Teil für das Verständnis der Arbeit. In zahlreichen Forschungsarbeiten wurde an Methoden zur Feuchtigkeitsmessung gearbeitet, die in verschiedenen Branchen wie Bauwesen, Lebensmittelindustrie oder Holzspeicherung eingesetzt werden. Diese Messverfahren können in direkte- und indirekte Messverfahren eingeteilt werden. Um auch in der Industrie genutzt zu werden, müssen diese Messverfahren bestimmte Eigenschaften aufweisen (Kupfer, 1997). Diese wären:

- zerstörungsfrei
- kontinuierlich
- zuverlässig
- unempfindlich gegen raue Industriebedingungen
- der Materialkontakt darf nicht entscheidend für den Messefekt sein (Kupfer, 1997)

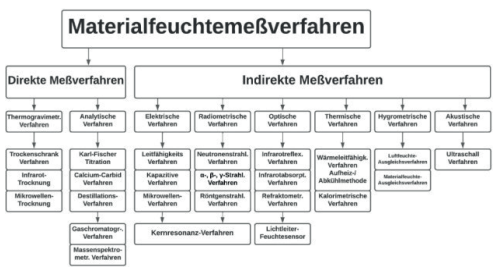


Abb. 1: Klassifizierung der Materialfeuchtemessverfahren (Quelle: In Anlehnung an *(Kupfer, 1997)*)

Abb. *1* zeigt unterschiedliche Materialfeuchtemessverfahren und deren Einteilung in direkte und indirekte Verfahren. Die Darr-Methode (Trockenschrank-Verfahren) und die CM-Methode (Calcium-Carbid-Verfahren) sind die bekanntesten direkten Verfahren, jedoch müssen Proben des zu messenden Materials vom Messmaterial entnommen werden. Die indirekten Verfahren sind zerstörungsfrei, aber sie können die Feuchtigkeit der Baustoffe nicht direkt messen. Stattdessen wird ein von der Feuchtigkeit abgeleiteter physikalischer Wert im Inneren eines bestimmten Volumens des zu messenden Baustoffes bestimmt.

Das kapazitive Messverfahren ist für eine zerstörungsfreie und kontaktlose Messung von Feuchtigkeit geeignet, da es keine Schäden am Material verursacht und keinen direkten Kontakt mit dem Material benötigt. Der entwickelte Feuchtesensor wird daher auf den theoretischen Konzepten des kapazitiven Verfahrens basieren.

Das kapazitive Verfahren bestimmt die Dielektrizitätskonstante mithilfe eines Hochfrequenzfeldes in einem Kondensator. Die Dielektrizitätskonstante von Wasser ist etwa 80 mal größer als die der Luft und mindestens 10 mal größer als die von mineralischen Baustoffen. In Tab. *1* werden Dielektrizitätskonstanten von Wasser und verschiedenen Baustoffen gezeigt.

Stoff	ε_r
Vakuum	1
Luft	1,006
Wasser	81
Porenbeton	2,041
Ziegel	5,304
Kalksandstein	3,984

Tab. 1 Dielektrizitätszahlen (Quelle: In Anlehnung an *(Leschnik & Schlemm, 1999)*)

Ein Kondensator ist ein elektrisches Bauelement, das aus zwei leitenden Flächen (Elektroden) besteht, die durch eine isolierende Schicht (Dielektrikum) getrennt sind. Durch das Anlegen einer Gleichspannung an die Elektroden werden Elektronen von einer Platte auf die andere übertragen, wodurch sich ein Ladungsunterschied zwischen den Platten ergibt.

Die Dielektrizitätskonstante ist ein Maß für die Polarisierbarkeit eines Dielektrikums. Sie gibt an, wie stark das Dielektrikum elektrische Felder abschirmt und wird meist mit dem griechischen Buchstaben Epsilon (ε) angegeben (Stiny, 2018). Im Fall eines Plattenkondensators wird die Dielektrizitätszahl in den folgenden Ausdruck einbezogen, um die Kapazität des Kondensators näherungsweise zu beschreiben:

$$C = \varepsilon_0 \cdot \varepsilon_r \cdot \frac{A}{d} \quad (1)$$

Wobei C die Kapazität des Kondensators in Farad (F), ε_0 die absolute Dielektrizitätskonstante, ε_r die relative Dielektrizitätskonstante, A die Fläche der Elektroden in Quadratmetern (m^2) und d den Abstand zwischen den Elektroden in Metern (m) beschreibt.

Das elektrische Feld, welches durch den Baustoff verläuft, sorgt für eine Änderung der Kapazität, wodurch man auf den Feuchtigkeitswert zurückschließen kann. Formel (1) gilt nur für den idealen Plattenkondensator. Für komplexere Geometrien müssen entsprechende numerische Simulationen eingesetzt werden.

Für die Simulation und zur Visualisierung der Funktionsweise der geplanten kapazitiven Messung wurde das Simulationsprogramm Comsol Multiphysics verwendet.

Für die Simulation wird die elektrische Energie E_{el} in Abhängigkeit von der Dielektrizitätskonstante des Baustoffes aufgetragen. Abb. *2* zeigt, dass die elektrische Energie mit steigender Dielektrizitätszahl größer wird.

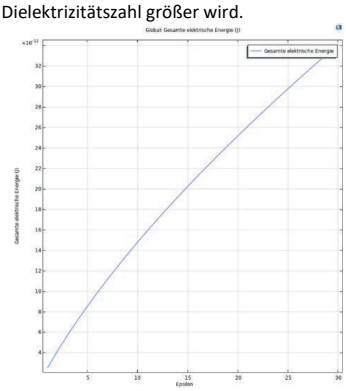


Abb. 2: Die elektrische Energie in Abhängigkeit von der Dielektrizitätskonstante

Der Zusammenhang zwischen der gespeicherten elektrischen Energie eines Kondensators und der Kapazität lautet:

$$E_{el} = \frac{1}{2} \cdot Q \cdot U \quad (2)$$

Der Zusammenhang mit der Kapazität lautet

$$C = 2 \cdot \frac{E_{el}}{U^2} \quad (3)$$

und kann damit aus Abb. *2* dargestellten gespeicherten Energiewerten ermittelt werden.

Für die Durchführung der Farbmessung wurde ein optisches Verfahren verwendet, das auf der Reflektion von Licht auf der Oberfläche des zu messenden Materials basiert. Hierfür wurde eine LED als Lichtquelle und eine Photodiode zur Messung des reflektierten Lichtes eingesetzt. Die Intensität des reflektierten Lichtes ist abhängig von der Farbe des Messmaterials, wobei dunklere Oberflächen geringere Intensitäten und somit niedrigere Messwerte aufweisen.

4. PRAKTISCHE DURCHFÜHRUNG

Vor der Fertigstellung der Platine wurde die Funktionalität des Sensors mittels eines Prototyps auf einem Steckbrett überprüft. Diese Überprüfung ergab zufriedenstellende Messergebnisse, wodurch die Entwicklung einer finalen Platine und ihre Bestückung mit den erforderlichen Komponenten gestartet wurde. Die ersten experimentellen Überprüfungen wurden unter Laborbedingungen durchgeführt.

5. ERGEBNISSE

Das Prinzip der Feuchtigkeitsmessung wird in Abb. *3* gezeigt und beruht auf drei Schritten. Im ersten Schritt wird ein Puls generiert (grünes Signal). Danach wird ein Filter verwendet, der in der Messelektronik eine Dreieckspannung erzeugt (blaues Signal). Die dritte Komponente ist ein Spitzenwertdetektor (rotes Signal), welcher in Verbindung mit der Dreiecksspannung den Spitzenwert und somit eine Aussage

über die Änderung der Feuchtigkeit angibt.

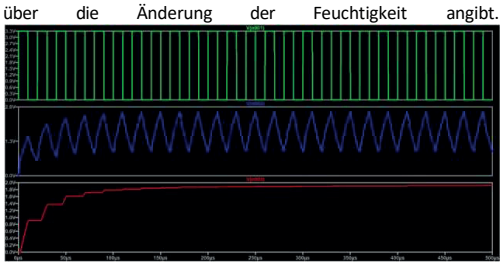


Abb. 3: Signale Feuchtigkeitsmessung

Die Messungen mit dem Prototypen zeigen ein ähnliches Verhalten wie die Simulation in Abb. *3*. Das Oszilloskop-Bild zeigt in Abb. *4* das entsprechende Rechtecksignal und den dazugehörigen Spitzenwert.



Abb. 4: Ausgangssignal Spitzenwertdetektor Schaltung

Die Messung erfolgt in drei Schritten. Zunächst wird die Elektrode in der Luft gemessen, anschließend in einem trockenen Messmaterial und danach im feuchten Material. Zum Abschluss werden die Elektroden aus dem Material entfernt und an der Luft erneut gemessen.

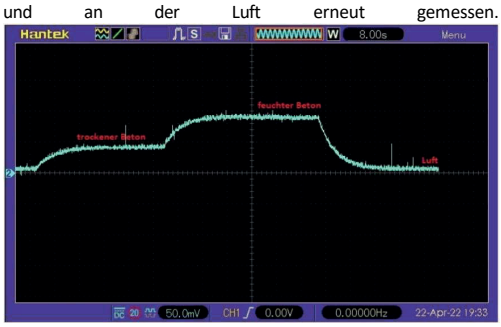


Abb. 5: kapazitive Messung der Feuchtigkeit

Die Messung in Abb. *5* zeigt einen klaren Unterschied zwischen Luft, trockenen und feuchten Messobjekt. Die Messwerte liegen bei feuchtem Material bei circa 80mV, bei trockenem circa 40mV und bei Luft bei nahezu 0V.

Die Auswertung der Photodiode erfolgt über den programmierten Web-Server. Dafür wurden ein weißes und ein eingefärbtes Material verwendet. Als Farbstoff wurde eine Rostlösung benutzt, welche die Farbe von Korrosion darstellen soll.



Abb. 6: Messung der unveränderten Farbe



Abb. 7: Messung der veränderten Farbe

Man sieht den Unterschied zwischen hellem und eingefärbtem Beton. Abb. *6* ergibt einen Messwert von zirka 0,7 mV bei hellem Beton. Eingefärbter Beton (Abb. *7*) weist einen Messwert von zirka 0,25mV für den eingefärbten beton auf.

6. ZUSAMMENFASSUNG UND AUSBLICK

Das Ziel dieser Arbeit war es, zu zeigen, dass es möglich ist, eine berührungslose Messung von Feuchtigkeit und Farbe an Oberflächen durchzuführen. Diese Technologie kann für die Messung von Feuchtigkeit und Rostbildung an Oberflächen in der Nähe von Befestigungspunkten eingesetzt werden. Recherchen und Tests haben ergeben, dass die kapazitive Messung die technisch sinnvollste Variante ist. Daraufhin wurden Messaufbauten entwickelt, um die Übereinstimmung von Theorie und Praxis zu gewährleisten. Vorabtests mit elektronischen Bauteilen haben gezeigt, dass die Einzelmodule funktionstüchtig sind, woraufhin ein entsprechender Prototyp entwickelt wurde. Die Software besteht aus Farb- und Feuchtigkeit sowie einem Webserver und einer Datenbank. Insgesamt hat die Arbeit gezeigt, dass die Entwicklung einer berührungslosen Messung von Feuchtigkeit und Farbe an Oberflächen möglich ist, jedoch gibt es noch einige Herausforderungen, die überwunden werden müssen, um genauere Ergebnisse zu erzielen. Weitere Forschungen sind erforderlich, um die Technologie weiter zu verbessern und die Messgenauigkeit zu erhöhen. Die Messung der Farbintensität gibt keinen Rückschluss auf die tatsächliche Farbe der Oberfläche, sondern lediglich auf den Farbunterschied. Die Feuchtigkeitsmessung zeigt einen klaren Unterschied zwischen trocken und feucht, jedoch kann keine Aussage über den absoluten Wert der Feuchtigkeit getroffen werden, da auf die Feuchtigkeit nur indirekt durch die Dielektrizitätskonstante geschlossen wird. Zudem bedarf es einer erneuten Kalibrierung des Sensors, wenn in einem anderen Material gemessen werden soll. Der nächste Schritt sollte eine Langzeitmessung an einem realen Versuchsaufbau sein, um den Einfluss von Feuchtigkeit und die Langzeitstabilität des Sensors genauer zu beobachten. Dafür sollte ein Testaufbau im Freien gewählt werden, wo Umwelteinflüsse wie Regen oder Schnee auch

auftreten können. Bisher wurde der Sensor unter Laborbedingungen getestet und es kann keine Aussage über Feldtests gemacht werden.

Literaturverzeichnis

Kupfer, K. (1997). *Editorial: Moderne verfahren zur materialfeuchtemessung.*

Leschnik, & Schlemm. (1999). *Dielektrische Untersuchung mineralischer Baustoffe in Abhängigkeit von Feuchte- und Salzgehalt bei 2,45 GHz.*

Stiny, L. (2018). *Grundwissen Elektrotechnik und Elektronik.* Springer Vieweg Wiesbaden.



Mustafa Oktay, B.Sc.

Attophotonics GmbH, 0262223495, mail@attophotonics.com

Berufliche Erfahrung
2021-2023 Attophotonics GmbH: Hardware- und Softwareentwicklung
2020-2021 Aviloo GmbH: Hardware- und Softwareentwicklung

Schulische Ausbildung
2019-2023 Masterstudium Mechatronik Fachhochschule Wiener Neustadt
Spezialisierung: Mechatronic Systems

2016-2019 Bachelorstudium Mechatronik Fachhochschule Wiener Neustadt
Spezialisierung: Micro & Nanoengineering

2010-2015 HTBLuVA Maschinenbau und Automatisierungstechnik

ENTWICKLUNG, SIMULATION UND OPTIMIERUNG EINES VIERBEINIGEN ROBOTERS

Sándor Pataky

Kurzfassung: *Ziel dieser Arbeit ist die Entwicklung eines mechanischen vierbeinigen Roboters, der sich auf unterschiedliche Arten bewegen kann. Der Roboter wird konstruiert und unterschiedliche Bewegungsarten werden simuliert und optimiert. Jede Bewegungsart ist statisch stabil, sodass der Bewegungsablauf jederzeit unterbrochen werden kann, ohne dass der Roboter umfällt. Damit ist eine sehr Der Roboter wird aber auch tatsächlich aufgebaut und bei Test gewonnene Ergebnisse genutzt, um die simulierte Bewegung möglichst realistisch zu machen, um die Gangarten zu verbessern. Die mechanischen Teile des Roboters werden 3D gedruckt, während die elektrischen Komponenten kommerziell verfügbar sind.*

Schlüsselwörter: *Bionik, vierbeinige Roboter, Simulation, statische Bewegung, Servo*

1. EINLEITUNG

Das Wort Roboter hat mehrere Bedeutungen und fasst eine große Menge von Geräten zusammen. Eine moderne Formulierung könnte lauten:
Eine Robot ist ein elektromechanisches Gerät, das nach vorheriger Programmierung unterschiedliche Prozesse oder Aufgaben erledigt. Diese Aufgaben / Tätigkeiten können durch direkten menschlichen Einfluss durchgeführt werden oder automatisiert mit Hilfe von Computerüberwachung.
Die ersten Roboter waren fixe, ortsgebundene Maschinen, heute können Roboter flexibel und ortsunabhängig sein.
Schon um 1980 gab es Interesse und Versuche, humanoide Roboter zu realisieren, wobei die Konstruktion, die Realisierung der Bewegung und die Regelung viel komplexer ist als bei Robotern die Räder besitzen. Die Fortbewegung ist komplizierter, da die Füße der Roboter nicht einfach ansteuerbar sind und Balance gehalten werden muss. Zwischen Füßen und Boden bestehet nicht immer Kontakt, wodurch sich Stabilität und Schwerpunk des Körpers permanent verändern. Schon bei der Bewegung auf einer Ebene ändert sich die Position des Schwerpunktes ständig. Um Kippen zu vermeiden, muss die Geschwindigkeit der Bewegung und die Schrittlängen betrachtet werden.
Honda untersuchte schon ab 1986 die zweibeinige Fortbewegung zu untersuchen. Die ersten erfolgreichen Ergebnisse wurden mit dem Modell E0 erreicht. Im Jahre 2000 wurde von Honda der Roboter ASIMO (Advanced Step In Innovative Mobility) entwickelt.
Heute ist Boston Dynamics eine der bekanntesten und erfolgreichsten Firmen der Roboter Entwicklung.

2. SERVO-TESTS

Modellbauservos sind die gewählten Antriebe der mechanischen Robotergelenke und spielen eine zentrale Rolle. Da die Modellbauservos keine industriellen Bauteile sind, die vom Hersteller getestet werden, werden sie dynamisch charakterisiert. Es müssen die Aspekte Genauigkeit, Geschwindigkeit, Widerholgenauigkeit sowie erzeugtes Drehmoment berücksichtigt werden. Die Auswertungen erfolgen dabei optisch mit einer Webcam
Die Auswertung erfolgt mit Matlab / Simulink.

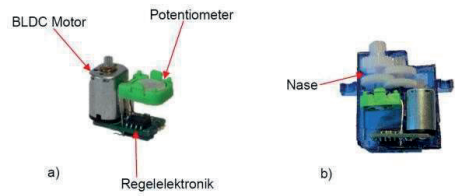


Abbildung 1 Servo Aufbau;
a) Servo mit Motor, Sensor und Elektronik ohne Gehäuse
b) Servo mit mechanischer Begrenzung (Nase)

Die meiste Servomotoren arbeiten im Bereich von 0° bis 180° und werden mechanisch die "Nasen" begrenzt, siehe Abb. 1. Der Gleichstrommotor (BLDC) kann kontinuierlich rotieren Es wird getestet, ob die Servos den theoretischen 180° Wert (je 90° in beide Drehrichtungen) wirklich erreichen. Die Servo-Positionen werden über PWM-Signale vorgegeben.

4.1 Bilderfassung

Während der Untersuchung ist auf dem Servohebel ist ein von der Kamera gut erkennbaren Zeiger montiert. Die Kamera ist über die Servo Drehachse parallel zur Zeigerebene montiert und erfasst die Servobewegung mit maximal 30 $Bilder/s$. Die Erfassungsrate hat einen Einfluss auf die Messgenauigkeit.
Der Zeiger wird jedem Bild automatisch erkannt und in einen Winkelwert umgerechnet. Nachdem jedem Zeitpunkt ein Winkel zugeordnet wird, können Geschwindigkeit ($\dot{\alpha}$) und Beschleunigung ($\ddot{\alpha}$) abgebildet werden.

4.2 Genauigkeit

Es werden Genauigkeit und Winkelbereich getestet. Jeder Test wird mit 3,3V und mit 5V Versorgung durchgeführt. Bei der Untersuchung von 5 Servos, bei 5V und 3,3V Versorgungsspannung zeigen sich im Winkelbereich Unterschiede von bis zu 5°.

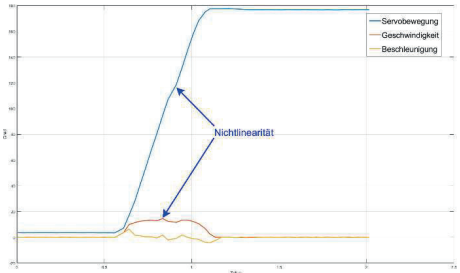


Abbildung 2 Servo Bewegung bei 3,3V Versorgungsspannung

4.3 Geschwindigkeit

Da der tatsächliche Drehbereich nicht mit dem Datenblattinformation übereinstimmt, wird auch die Geschwindigkeit der 5 Servos untersucht.

Die Bewegung des Servo kann man in drei Teile aufteilen: Beschleunigungsphase, lineare Bewegung und Abbremsphase.
Die Beschleunigungs-/Abbremsphasen (siehe Abb.2) sind einander ähnlich. Die Geschwindigkeit steigt maximal. Dann bewegt sich der Servo mit einer konstanten Geschwindigkeit bis zur Abbremsphase.
Während die ersten zwei Phasen bei allen getesteten Servos ähnlich verhalten, gibt es bei der Bremsphase deutliche Unterschiede.
Da es bei 5V Versorgung zu einem überschwingen kommt, so dass die höhere Versorgung zu einer verbesserten Dynamik führt (siehe Abb.3).

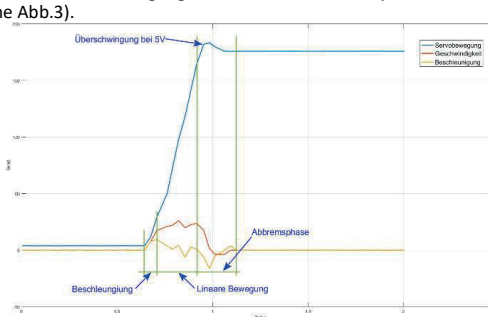


Abbildung 3 Servo-Bewegung Diagramm (bei 5V Versorgungsspannung) für Sollvorgabesprung von 0 auf 180°

4.4 Wiederholgenauigkeit und Belastung

Da die Servos im Rahmen der Vorwärtsbewegung periodische Bewegungen ausführen, ist es die Wiederholgenauigkeit wichtig.
Die Genauigkeit wird mit einem kontinuierlichen Sinussignal getestet.
Dabei werden die Tests mit und ohne Momentenbelastung durchgeführt. Ohne Belastung verhält sich der Servo wie erwartetet, mit Belastung stellt sich der in Abb. 4 gezeigte verlauf ein.

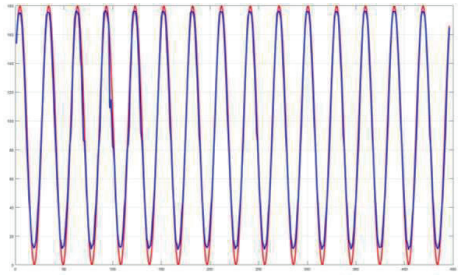


Abbildung 4 Vorgegebenes Signal (rot) und gemessenes Signal (blau) bei 1Hz (Servo unter Belastung)

5. STATISCHE UND DYNAMISCHE BEWEGUNGSARTEN

Bei der Fortbewegung ist es ein stabiler Gang wichtig. Eine stabile Bewegung kann auf zwei Arten erreicht werden: statisch oder dynamisch.

5.1 Statisch stabile Bewegung

Eine Bewegung ist statisch stabil, wenn sich „der Körperschwerpunk entlang der Richtung des Gravitationsvektors (auf dem Boden) innerhalb der durch drei oder mehr Ecken gebildeten konvexen Hülle der Bodenkontaktpunkte befindet“ [2]. Bei vier- oder mehrbeinigen Robotern entsteht eine statisch stabile Fortbewegung, wenn mindestens drei Füße direkt am Boden sind und der Roboter ohne extern einwirkende Kräfte nicht umkippt.

5.2 Dynamisch stabile Bewegung

Bei dynamisch stabilen Bewegungen mit weniger als drei Kontaktpunkten wird das Kontaktpolygon zu einer Linie, einem Punkt oder es verschwindet völlig. Während der Fortbewegung darf der Körperschwerpunkt außerhalb des Kontaktpolygons sein. Um ein Umfallen zu vermeiden, muss sich der Körperschwerpunkt relativ zu den Kontaktpunkten ändern, etwa beim Stehen. Alternativ können auch die Kontaktpunkte neu gewählt werden, bevor eine instabile Situation eintritt. Dies ist etwa beim Gehen der Fall.
Eine spezielle Art des dynamischen Gehens ist dynamisches Laufen. bei dem zeitweise zwischen Boden und den Füßen kein Kontakt entsteht, dh. der gesamte Körper in der Luft ist. Diese Bewegung bedeutet nicht unbedingt, dass sich der Roboter nur nach vorne bewegt
Obwohl beide (statische und dynamische) Bewegungsarten als stabil bezeichnet werden, zeigt sich der große Unterschied, wenn die Bewegung plötzlich endet. Statisch stabil bewegte Körper fallen nicht um, während dynamisch stabile Körper meist umfallen.

5.2.1 Bewegungsarten von vierbeinigen Lebenswesen und Roboter

Vierbeinige Lebenswesen werden in zwei große Gruppen aufgeteilt: Säugetiere und Reptilien. Beim Körperaufbau der Säugetiere sind die Beine meistens unter dem Körper, siehe Abb.5, während die Beine bei Reptilien seitlich weggestreckt sind. Reptilien haben dadurch ihren Körperschwerpunkt näher beim Boden und bewegen sich sehr stabil. Einen weiteren Unterschied liegt im Bewegungsablauf, der die Geschwindigkeit bestimmt.



Abbildung 5 Typisches Säugetier (Hund) Aufbau mit Beinen unter dem Körper während Bewegung

5.2.1.1 Trott Bewegung

Bei Trotten führen die diagonal liegenden Beinpaare gleichen Bewegungen aus, wodurch eine schnellere Fortbewegung als das Gehen erreicht wird. Von Seitenansicht macht der Körper eine schwingende Bewegung Diese Bewegung gehört zu den statisch dynamischen Bewegungen. Abhängig von der Geschwindigkeit und der Auslenkung der Massenschwerpunk, kann es sein, dass alle vier Beine am Boden sind. Wenn die Bewegung schneller ist, dann kann es vorkommen, dass keine Beine mit dem Boden gar nicht in Berührung kommen, was jedoch noch nicht als „Laufen" bezeichnet wird. [3]

5.2.1.2 Sprungartige Bewegung

Diese Bewegungsart nutzen eher kleinere Tiere wie Hunde, Katzen oder Eichörchen. Bei dieser Bewegung arbeiten die Beine in Paaren. Diese Bewegungsart gehört schon zum Laufen Zur Fortbewegung drückt das hintere Beinpaar den Körper nicht nur nach vorne, sondern gleichzeitig nach oben. Die Aufgabe der vorderen Beine ist die Dämpfung der Landung und das Hochheben des Körpers.

5.2.1.3 Krabbelnde Bewegung

Krabbeln ist dadurch gekennzeichnet, dass immer nur einen Fuß in der Luft ist (siehe Abb.6). Es ermöglicht für vierbeinige Roboter eine langsame aber statisch stabile Bewegung. Während der Fortbewegung ändert sich die Lage des Schwerpunkts des Körpers ständig.
Für den Roboter dieser Arbeit wird eine krabbelnde Bewegung gewählt. Für andere Gangarten wären Servomotoren als Antriebe nicht geeignet. Darüber hinaus ist die Implementierung einer krabbelnden Bewegung günstig, weil dabei die Geschwindigkeit des Ablaufes keinen Einfluss auf die Stabilität der Bewegung hat. Roboter, die einen dynamischen Gang nutzen sind neben den hohen Anforderungen an den Aktuator auch aus regelungstechnischer Sicht sehr komplex und anspruchsvoll

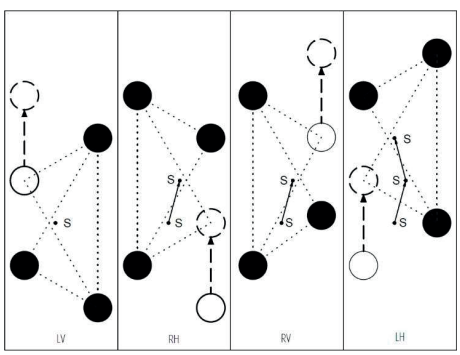


Abbildung 6 Krabbel Roboter: die vier Phasen der statischen Bewegung-auch die Schwerpunktstrajektorie ist eingezeichnet

6. KONSTRUKTION UND AUFBAU DES ROBOTERS

6.1 Allgemeine Rahmenbedingungen

Da der Roboter möglichst kompakt sein soll, dürfen die Bauteilen auch nicht groß sein. Während bei größeren Robotern die Auswahl der ausreichend starken Aktoren weniger problematisch ist, ist sie bei kleineren Robotern eine Herausforderung. Je kleiner ein Roboter ist, desto schwieriger ist es den strukturellen Aufbau analog zu Tieren zu realisieren, und umso schwieriger ist es ausreichend kleine, aber starke Aktuatoren zu finden [4].
Der Roboter muss statisch stabil sein, weshalb mindestens 4 Beine notwendig sind. Da die Beine mindestens zwei Freiheitsgrade besitzen müssen, um eine flüssige Bewegung zu erzeugen sind mindesten 8 Aktuatoren erforderlich.
Als Aktoren und zur Steuerung werden kommerzielle Bauteile verwendet, die jederzeit einfach ersetzbar oder wiederherstellbar sind. Für die Bewegung der Beine werden deswegen Modellbau-Servos verwendet.
Aufgrund der Größe und des relativ geringen Preises, werden 3 Servos pro Bein eingesetzt, wodurch deutlich mehr Bewegungsmöglichkeiten möglich sind.

6.2 Auswahl der Steuereinheit

Bei der Auswahl des Mikrocontrollers spielen Speicherplatz, Versorgungsspannung, Anzahl und Art (digital, analog, I2C, usw.) der Anschlüsse, Flash Memory, SRAM, Abmessungen und der Preis eine wichtige Rolle. Die in Frage kommenden Mikrocontroller sind Arduino Uno, Arduino Mega, ESP8266 (V2/V3) und ESP32
Aus Sicht von Speicherplatz, Abmessungen, Preis und die extra Funktionen wäre die bevorzugte Variante der ESP32. Dieses Modell ist mit der Arduino IDE programmierbar und durch die große Anzahl von Anschlüssen ermöglicht es die Verwendung von mehr als 12 Servos und weiteren Bauteilen. Der größte Nachteil der ESP-Serie ist jedoch, dass sie mit Matlab und Simulink nicht programmierbar sind. Im Falle einer direkten händischen Programmierung kann ein ESP- Mikrocontroller eine gute Alternative sein, aber nachdem Simulation und Programmierung in Matlab / Simulink ausgeführt werden wird ein Arduino bevorzugt [5]. Da für die Beinbewegungen 12 Servos erforderlich sind. Vorteil des Arduino Mega ist die Anzahl (16) der I/O Pins die PWM-Signale ausgeben können. Daher wird kein zusätzlicher PWM-Treiber benötigt. Weitere Komponenten bedeuten zusätzlichen Platz und weitere Einschränkungen, denn bei elektrischen Bauteilen sind Kabel und Kabelanschlüsse immer auch Schwachstellen.

6.3 Größe, Geometrie und Herstellverfahren

6.3.1 Aufbau des Körpers

Damit der Servo nur kleine Drehmomente bereitstellen muss, werden die Beinteile möglichst kurz und stabil konstruiert. Bei der Konstruktion muss beachtet werden, dass die Teile durch 3D- Druck hergestellt werden. Diese Technik ist billig, präzise, erlaubt komplexe Geometrien und die Komponenten können gewichtoptimiert hergestellt werden. Die Teile werden mit einen FDM (Fused Deposition Modelling) 3D Drucker aus PLA hergestellt. Bei der Materialauswahl wird PLA (Poly Lactic Acid) gewählt, da es einfach zu drucken ist, und eine hohe Festigkeit und eine ausreichende UV-Beständigkeit aufweist.
Den Körper kann man auf zwei große Gruppe aufteilen, Grundkörper und Beine.

6.3.2 Grundkörper

Der Grundkörper besteht aus drei Teilen, wobei nur zwei Teile (obere- und untere Gehäuseteil) einen Einfluss auf die Bewegung und die Festigkeit der Körper haben, während die Elektronik Abdeckung eher sicherheitstechnische- und designtechnische Funktionen hat, siehe Abb.10. Alle Körperteile werden mit einer Gitterstruktur konstruiert, um Material und Gewicht zu sparen. Diese Gitterstruktur ermöglicht eine einfache schnelle Befestigung von extra Teilen (z.B.: Arduino, Verteiler, Akku). Auf dem oberen Körperteil werden die ersten Servomotoren festgeschraubt, während an dem unteren Teil die Drehstützen zur Positionierung verwendet werden.
Der Arduino findet innerhalb des Körpers Platz, auch der für die Anschlüsse erforderliche Servo- Verteiler.
Arduino, Verteiler und die sind über eine Abdeckung zugänglich.

6.3.3 Beine

Der Roboter hat zwei Beinpaare, die symmetrisch aufgebaut sind. Jeder Servo wird an einem Beinsegment mit je zwei Schrauben befestigt. Zur Bewegung der Beine werden die Servohebel fest in Stellen mit einer feinen Passung eingesetzt, wodurch keine extra Befestigung der Servohebel notwendig ist. Die Servohebel werden die mit den Servos verschraubt und der Abstand der Beinteile ist fix eingestellt, siehe Abb. 7.

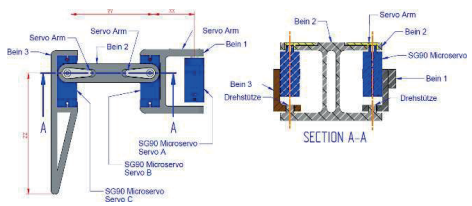


Abbildung 7 | Bein Aufbau

Unter jedem Servo befindet sich zur Ausrichtung der Servo-Achse eine Drehstütze, um eine saubere Bewegung zu ermöglichen

7. SIMULATION

Mit dem virtuellen Modell können Test durchgeführt werden, die beim realen Roboter schwer realisierbar wären oder große Kosten verursachen. Bei einem digitalen Modell können an schwer zugänglichen Stellen oder sogar in einem Bauteil Analysen durchgeführt werden, wie zB. Ermittlung von Temperaturverteilungen, Deformationen oder die Messung von Momenten (in den Servos), Kräften (Aufstandspunkt), Drehraten oder Beschleunigungen. Da die Analysen und deren Ergebnisse am Computer vorliegen, können sie analysiert, wiederholt, erweitert oder korrigiert wurden
Zu Simulation, wird ein CAD Modell benötigt. Während der Konstruktion wird kontrolliert ob zwischen den Komponenten Kollisionen auftreten. Simulationen und virtuellen Tests werden in Simulink Multibody ausgeführt. Simulink bietet die Möglichkeit, einfache Körper direkt zu erstellen, oder fertige CAD-Modelle zu importieren.
Bei der Modellerstellung wird zuerst der Grundkörper definiert. Da die Beine ähnliche Bewegungen ausführen, wird zuerst ein Bein optimal eingestellt und die weiteren Beine durch Kopieren und Anpassungen ergänzt. Bei der Verbindung zwischen den Körpern über die Servos wird angenommen, dass die Drehbewegung (Derhwinkel) vorgegeben sind. Das Modell bestimmt die für eine Bewegung erforderliche Drehmomente.
Für die Simulation der Bewegung eines einzelnen Beines ist eine kinematische Analyse der Bewegung erforderlich. In [6] ist ein Matlab Code zur Beschreibung der inversen Kinematik eines Beines vorgestellt, das im Aufbau jenem des Roboters entspricht. Ausgehend vom Befestigungspunkt des Beines, siehe Abb.8, können damit die Gelenkswinkel für einen gewünschten Punkt der Beinspitze „A“ berechnet werden.

Sind die Bewegungspunkte von „A“ relativ zum Koordinatensystem (X, Y, Z) gegeben, siehe Abb 8., können daraus die erforderlichen Gelenkwinkel $\varphi A(t)$, $\varphi B(t)$ und $\varphi C(t)$ als Funktion der Zeit und gefunden werden
Für die Simulation werden noch Boden, die Gravitation vorgegeben. dann bleibt der Vierfüßer am Boden. Wenn zwischen den Beinspitzen A und Boden in der Form eines Feder-Dämpfer Element realisiert sind, die entsprechend der Kontaktbedingungen an- oder abgeschaltet wird. Daneben ist es aber auch für die Bereitstellung der horizontalen Reibkräfte verantwortlich, so dass sich der Roboter vorwärts bewegen kann. Die Modellierung als Reibelement erlaubt auch die Berechnung der Fortbewegung bei kinematisch nicht eindeutiger Bewegung der Roboterbeine.

In der Simulation werden die Drehwinkel $\varphi 1(t)$, $\varphi 2(t)$, $\varphi 3(t)$ der Servos vorgegeben und daraus die erforderlichen Drehmomente $MA(t)$, $MB(t)$ und $MC(t)$ berechnet. Um die Servo Momente zu bestimmen, muss die Massenverteilung der Komponenten bekannt sein. Diese können manuell angegeben, oder aus Geometrie errechnet werden. Dabei ist im Eigengewicht der Komponenten der Akku, der Controller etc. nicht enthalten. Daher wird angenommen, dass diese Elemente zusammen ein Gewicht von m = 0,130 − 0,350kg besitzen, welche zusätzliche Masse ergänzt wird.

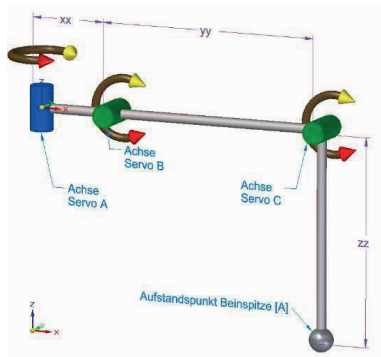


Abbildung 8 Definition der Gelenkwinkel eines Beines; blaues Gelenk drehen sich um die Z-Achse und grüne Gelenke um die Y-Achse.

Nachdem eine konstante Vorwärtsgeschwindigkeit erwünscht ist, sind für Servo A kaum Momente erforderlich. Die größten Momente bringt Servo B auf, da er das Robotergewicht kompensieren muss, und einen relativ großen Hebelsarm besitzt. Servo C hingegen ist für die Position des Kontaktpunktes verantwortlich und besitzt nur geringe Momente, siehe Abb.9.



Abbildung 9 Servo Momente für ein Bein Krabbelbewegung, Es werden vier Schrittfolgen durchlaufen

7.1 Bewegungstypen Simulation

Es ist notwendig mehrere unterschiedliche Arten von Bewegungen zu definieren, um auf unterschiedliche Situationen zu reagieren. Da sich der Roboter im Rahmen der Arbeit nur auf ebenem Boden bewegt, ist die Anzahl der Bewegungen geringer und es sind keine Sensoren erforderlich, um den Kontakt zum Boden zu detektieren.
Bei ebenem Untergrund sind folgende Bewegungsarten erforderlich: geradlinige und seitliche Bewegung, Bewegung auf einem Kreisbogen mit variablem Radius oder Umdrehen an einer Stelle erforderlich.

7.1.1 Normaler Reptiliengang

Die einfachste Art, um eine geradlinige Bewegung zu realisieren, besteht darin, den Aufstandspunkt relativ zum Roboter entlang der Bewegungsrichtung zu verschieben.
Bei Simulation der Fortbewegung in Simulink, erscheint die Bewegung optimal. Der Roboter bewegt sich stabil und kippt nicht um.

7.1.2 Umdrehen am Stand

Eine Möglichkeit die Richtung zu ändern ist die Drehung am Stand. Dabei wird die Position des Schwerpunktes im Raum nicht geändert. Bei symmetrischem Aufbau mit gleichen Abständen der Drehachsen ist diese Bewegung einfach realisierbar. Theoretisch sollten die Aufstandspunkte eine Kreisbahn beschreiben. In der Simulation wird diese durch die Tangenten approximiert, sodass die Phasen der Beinbewegung wieder durch ein Rechteck beschrieben werden, und analog zu geradliniger Bewegung sind.
Aus diesem Grund wird beim Umdrehen am Stand eine Schwerpunktverlagerung durchgeführt, die dadurch beschrieben wird, dass sich der Schwerpunkt des Roboters auf einer überlagerten kleinen Kreisbahn bewegt. Die Überlagerung garantiert, dass sich der Roboterschwerpunkt beim Umdrehen im Kontaktpolygon befindet.

7.1.3 Bewegung am Kreisbogen

Um einen beliebigen Endpunkt zu erreichen, reicht die geradlinige Bewegung in Kombination mit der Drehbewegung. Jedoch kann die Durchschnittsgeschwindigkeit durch eine kontinuierliche Bewegung auf einer Kreisbahn erhöht werden. Die einfachste Art eine Kurve zu machen, besteht darin, die Schrittlängen an einer Seite kürzer zu wählen. Auch wenn diese Wahl kinematisch nicht perfekt ist, und teilweise zu einem „Rutschen“ im Aufstandspunkt führt, ist sie sehr einfach zu implementieren, weil die Funktion des geradlinigen Ganges nur leicht angepasst werden muss. Die Größe (Radius) der Kreisbahn ist durch die Schrittlängen auf beide Seiten, die Anzahl der Beine und der Aufbau des Roboters definiert. Bei zu langen oder zu kurzen Schritten kann der Roboter einen instabilen Zustand erreichen und kippen. Diese Instabilität wird sowohl bei der Simulation als auch während des Tests erkannt, und durch die Wahl der Schrittlänge korrigiert.

7.2 Reale Tests

Nachdem die unterschiedlichen Bewegungsarten erfolgreich simuliert wurden, werden die Ansteuersignale als Code auf die Steuereinheit des Roboters geladen, um echte Tests durchzuführen. Bei diesen sind mechanische und softwaretechnische Probleme aufgetaucht, die ausgewertet, und korrigiert werden.
Zunächst wird der Roboter am Körper gehalten und die Bewegung der Beine kontrolliert. Auch die Bewegungsabfolge der Beine in richtiger Reihenfolge wird geprüft. Wenn die Bewegung der Beine optimal erscheint, wird der Vierfüßer auf den Boden gestellt und weitere Tests durchgeführt.

7.2.1 Normaler Reptiliengang

Im ersten Test wird der Code von normalem Reptiliengang auf den Arduino hochgeladen.
Das Problem bei der realen Bewegung ist die Lage des Schwerpunktes, der sich nicht innerhalb des von Beinen generierten Stabilitätsdreiecks befindet.
Daher wird der Körper um $50mm$ verlängert. So wird der Schwerpunkt weiterhin in der Mitte des Aufbaus gehalten, es entsteht aber eine bessere Gewichtsverteilung und mehr Arbeitsraum für die Beine. Die wackelige Bewegung wird damit reduziert, die Fortbewegung wird sauberer und der Roboter fällt nicht um. Zusätzlich wird der Code modifiziert. Beim Aufheben eines Beines wird das Gewicht in Richtung des gegenüberliegenden Beines verlagert, sodass der Schwerpunkt im Stabilitätsdreieck bleibt.

7.2.2 Umdrehen am Stand

Während des Umdrehens ist der Roboter stabil und hält seine Position gut. Wenn der Roboter allerdings nicht mit einem Akku versorgt wird, kann das Versorgungskabel die Bewegung behindern und muss von oben gehalten werden

7.3 Auswertung und nächste Schritte

Die Bewegungen sind nach einigen Feineinstellungen stabil, jedoch gibt es Potential zur weiteren Optimierung der Bewegungen. Die stabilste Bewegung ist das Seitenkrabbeln, wobei sich der Roboter nach links oder rechts bewegt oder mit seiner Bewegung eine Kreisbahn beschreibt. Sobald der Radius des Kreises ausreichend groß ist, ist diese Bewegung stabil und zu bevorzugen.
Die Bewegung des Vierfüßers wird auf unterschiedlichen Oberflächen getestet. Auf Grund zu geringer Reibung zwischen Fuß und Boden werden Gummihüllen als "Schuhe" am Roboter fixiert, siehe Abb.10.

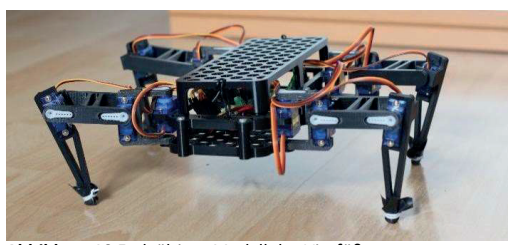


Abbildung 10 Endgültiges Modell des Vierfüßers

9 ZUSAMMENFASSUNG

Im Rahmen dieser Arbeit wird ein mechanischer vierbeiniger Roboter entwickelt, der unterschiedliche Gangarten und damit unterschiedliche Bewegungen ausführen kann. Die Arbeit beschreibt den Aufbau, die Komponenten und deren Tests, die Hardware und softwaretechnische Steuereinheiten, die Simulationen und die Tests mit dem Roboter.

Zunächst werden die Bewegungen mit Hilfe von kinematischen Analysen entwickelt und mit Hilfe einer Mehrkörpersimulation numerisch analysiert. Dabei liefern die Simulationen bereits wichtige Informationen in Bezug auf Stabilität und kinematisch korrekte Bewegungen. Trotzdem wird von realen Bedingungen ausgegangen, sodass erst reale Tests weitere Informationen liefern.

Der größte und bedeutendste Unterschied zwischen Simulation und Test zeigte sich bei der Stabilität. Während bei der Simulation die Bewegung in meisten Fällen stabil wirkt, wird beim Test öfter ein grenzstabiler oder instabiler Zustand erreicht. Um diese Instabilität zu beheben, wurden die Geometrie und Abmessungen des Körpers geändert.

Eine weiterer Einflussfaktor für unzureichende und zeitweise instabile Bewegungen, sind die Servos. Die billige kommerzielle Komponenten sind für eine kontinuierliche und präzise Ansteuerung zu ungenau, weshalb hochwertige Servos eingebaut werden müssen.

Der in der Arbeit entwickelte vierbeinige Roboter kann bereits mehrere Fortbewegungsarten ausführen und beherrscht in der Ebene den Vorwärtsgang, Umdrehen am Stand, den Kurvengang vorwärts sowie den Gang seitlich (Krabbengang) und den Gang seitlich mit Kurven. Alle Bewegungen des Roboters sind statisch stabil, sodass der Roboter jede Bewegung unterrechen oder sehr langsam ausführen kann. Für die Bewegung sind keine Kontaktsensoren erforderlich. Die Programmierung der Arduino Steuereinheit erfolgt mit automatischer Code- Generierung direkt aus der Simulink-Simulationsumgebung. Die Arbeit zeigt, dass mit Hilfe von 3D Druck und kommerziellen Modellbau- Servoantrieben ein vierbeiniger Roboter entwickelt werden kann, der sich zuverlässig und sicher bewegt. Auf diese Arbeit aufbauend können zukünftig Bewegungen auf unebener Umgebung sowie die Erfassung der Umgebung mit Hilfe von Sensoren durchgeführt werden.

10. LITERATURVERZEICHNIS

[1] J. Pearce, „The New York Times,“ 15 08 2011. [Online]. Available: https://www.nytimes.com/2011/08/16/business/george-devol-developer-of-robot-arm-dies-at-99.html. [Zugriff am 07 07 2022].

[2] S. S. Hardt, „Modellierung und Simulation der Dynamik des Laufens bei Roboter,“ *TUD Forschung,* 2002.

[3] A. K. Mishra, *Design, Simulation, Fabrication and Planning of Bio-Inspired Quadruped Robot,* Indian Institute of Technology Patany, 2014, p. 54.

[4] Siciliano and Khatib, in *Handbook of Robotics*, 2nd ed., Springer-Verlag, 2016, p. 545.

[5] Matlab & Simulink - Embeded Coder User's Guide, Matlab, 2022.

[6] P. Corke, “Robotics Vision and Control,” 1st ed., Springer-Verlag, 2016, pp. 221-225.

[7] Angermann, Beuschel und Rau, „Matlab-Simulink-Stateflow,“ 6. Auflage Hrsg., München, Oldenbourg, 2009, p. 427.



BSc. Sándor Pataky

Fachhochschule Wiener Neustadt, Johannes-Gutenberg-Straße 3, 2700 Wiener Neustadt, +43 677 6176 1221, psanyesz95@gmail.com

Ich bin im Jahr 1995 in Budapest, Ungarn geboren. Im Jahr 2015 bin ich nach Österreich gezogen und and der Fachhochschule Wienerneustadt die Bachelor- und Masterstudium (Mechatronik) gemacht. Nach dem Bachelor- neben des Masterstudium habe ich an der Fachhochschule als Studienassistenz und bei der JCT Analysentechnik als Konstrukteur gearbeitet.

PARKSPERR MODELLIERUNG UND ECHTZEITSIMULATION

Petsch Philipp

Kurzfassung: *Diese Arbeit befasst sich mit der Modellierung eines generischen Parksperrmodells in der Matlab/Simulink Umgebung für einen HIL-Prüfstand. Ziel ist es ein Modell zu entwerfen, welches dem Verhalten einer realen Parksperre möglichst nahe ist. Das ist erforderlich, damit das Verhalten eines Steuergeräts nicht von einem Fehlverhalten des Parksperrmodells beeinflusst wird.*
Das Modell wird in mehrere Subkomponenten unterteilt, die lauten BLDC-Motor, Inverter, Parksperrmechanik und Sensorik. Die Modellierung erfolgt abstrakt durch mathematische Formulierungen. Abschließend werden drei unterschiedliche Testversuche durchgeführt, bei denen das Modell mit der realen Parksperre verglichen und validiert wurde. Die Ergebnisse zeigen, dass das Parksperrmodell im HIL-Prüfstand die reale Parksperre ersetzen kann.
Schlüsselwörter: *HIL, Parksperre, ECU, Matlab/Simulink*

1. EINLEITUNG

HIL-Prüfstände spielen eine enorm wichtige Rolle in der Entwicklung im Automotive Bereich. Damit Funktionen einer ECU, mit Parksperre, in einem frühen Entwicklungsstadium getestet werden können, ist es notwendig eine reale Parksperre in den HIL-Prüfstand zu implementieren. Zusätzlich sind im Unternehmen HIL-Prüfstände mit realen Parksperren begrenzt. Durch die Entwicklung eines Parksperrmodells soll die reale Parksperre ersetzt werden. Ein generisches Parksperrmodell lässt sich ohne Probleme in eine HIL-Simulation integrieren und durch Anpassung der Modellparameter können projektspezifische Eigenschaften der Parksperre eingestellt werden.

In seltenen Fällen werden heutzutage noch Prüfläufe regelmäßig manuell durchgeführt. In der Regel erfolgt dies bereits automatisiert. Das Entwickeln, Testen und Debuggen dieser Automatisierungsskripte sollte aber nicht am kostenintensiven Prüfstand mit echten Leistungskomponenten geschehen. HIL-Prüfstände bieten die Möglichkeit komplette Systeme virtuell zu simulieren und somit kostengünstig Testabläufe zu simulieren.
Um die Ziele dieser Arbeit zu erreichen werden Literaturrecherchen, gleichende Studien und Versuche durchgeführt.

2. PROBLEM- UND AUFGABENSTELLUNG

Der Wettbewerb in der Automobilbranche ist groß und um wettbewerbsfähig zu bleiben müssen günstige Entwicklungsmethoden gefunden werden. Auf lange Sicht sind virtuelle Simulationen und virtuelle Verifikation ressourcenschonend. Aus diesem Grund sollen Systemkomponenten eines Antriebstranges mit möglichst realistischen Systemverhalten als Modell nachgebildet werden. Diese Arbeit befasst sich mit der Parksperre, einem mechatronischen System, das elektrisch und mechanisch beschrieben werden soll. Die Komponenteneigenschaften sind beim elektrischen Modell bekannt, die des mechanischen Modells werden durch Probieren angenommen. Die Aufgabe besteht darin ein zuverlässiges Modell zu entwickeln, das dem realen Systemverhalten sehr nahe ist. Im HIL-Prüfstand soll das Testobjekt, ein Steuergerät, den Unterschied zwischen realer und modellierter Parksperre nicht erkennen.

3. SIMULATIONSMODELL

Die Arbeit befasst sich zu Beginn mit der Modellierung des BLDC-Motors. Als Grundlage für die mathematische Beschreibung des Motors diente eine weitere Masterarbeit, welche die Modellierung und Steuerung eines BLDC-Motors beschreibt [1]. Die Basis für das elektrische Modell bilden die drei „Line to Line“ Spannungsgleichungen v_{ab}, v_{bc} und v_{ca},

$$v_{ab} = R(i_a - i_b) + L\frac{d}{dt}(i_a - i_b) + e_a - e_b \tag{1}$$

$$v_{bc} = R(i_b - i_c) + L\frac{d}{dt}(i_b - i_c) + e_b - e_c \tag{2}$$

$$v_{ca} = R(i_c - i_a) + L\frac{d}{dt}(i_c - i_a) + e_c - e_a \tag{3}$$

wobei R den elektrischen Widerstand, i den Phasenstrom, L die Induktivität und e die Back EMF Spannung beschreibt.

Die Back-EMF Spannungen e_a, e_b und e_c sind das Produkt aus Maschinenkonstante k_e, Winkelgeschwindigkeit ω_m und den Funktionsverlauf $F(\theta_e)$ der Back EMF.

$$e_a = k_e\omega_m F(\theta_e) \tag{4}$$

$$e_b = k_e\omega_m F\left(\theta_e - \frac{2\pi}{3}\right) \tag{5}$$

$$e_c = k_e\omega_m F\left(\theta_e - \frac{4\pi}{3}\right) \tag{6}$$

Die Drehmomentengleichung T_e,

$$T_e = \frac{k_t}{2}\left[F(\theta_e)i_a + F\left(\theta_e - \frac{2\pi}{3}\right)i_b + F\left(\theta_e - \frac{4\pi}{3}\right)i_c\right] \tag{7}$$

berechnet sich aus der Drehmomentkonstante k_t und den mit $F(\theta_e)$ gewichteten Phasenströmen. Wobei F den Funktionsverlauf der Back EMF und θ_e die elektrische Winkelposition beschreibt. Da beim BLDC-Motor die Phasen im Sternpunkt verknüpft sind, kann mit Hilfe der Knotenbedingung die Glg. 1-3 reduziert werden.

$$i_a + i_b + i_c = 0 \tag{8}$$

Daraus folgt

$$v_{ab} = R(i_a - i_b) + L\frac{d}{dt}(i_a - i_b) + e_a - e_b \tag{9}$$

$$v_{bc} = R(i_a + 2i_b) + L\frac{d}{dt}(i_a + 2i_b) + e_b - e_c \tag{10}$$

Mit Hilfe des mechanischen Drallsatzes $\dot{\omega}_m J = \sum M$ sowie $\dot{\theta}_m = \omega_m$ folgt mit $M = T_e - T_L$ das Zustandsraummodell, siehe Glg. 11

$$\begin{pmatrix} i'_a \\ i'_b \\ \omega'_m \\ \theta'_m \end{pmatrix} = \begin{pmatrix} -\frac{R}{L} & 0 & 0 & 0 \\ 0 & -\frac{R}{L} & 0 & 0 \\ 0 & 0 & -\frac{k_f}{J} & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} i_a \\ i_b \\ \omega_m \\ \theta_m \end{pmatrix} + \begin{pmatrix} \frac{2}{3L} & \frac{1}{3L} & 0 \\ -\frac{1}{3L} & \frac{1}{3L} & 0 \\ 0 & 0 & \frac{1}{J} \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} v_{ab} - e_{ab} \\ v_{bc} - e_{bc} \\ T_e - T_L \end{pmatrix} \quad (11)$$

mit der Ausgangsgleichung

$$\begin{pmatrix} i_a \\ i_b \\ i_c \\ \omega_m \\ \theta_m \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ -1 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} i_a \\ i_b \\ \omega_m \\ \theta_m \end{pmatrix} \quad (12)$$

Dabei beschreiben T_e, T_L und J das elektrisch erzeugte Moment, das mechanische Lastmoment und das Massenträgheitsmoment von Rotor und Getriebe.
Eine Summe aus Drehmomenten beschreibt das mechanische Modell, das gegen das elektrische Drehmoment wirkt.

$$T_{Diff} = T_e - T_{Reibung} - T_{StopperLinks} - T_{StopperRechts} - T_{Feder} \quad (13)$$

Die Momentendifferenz T_{Diff} berechnet sich aus dem elektrischen Drehmoment T_e, dem Reibungsmoment, dem mechanischen Anschlag links und rechts mit $T_{StopperLinks}$ und $T_{StopperRechts}$ und dem Drehmoment der Torsionsfeder T_{Feder}.
Das Drehmoment der Torsionsfeder berechnet sich aus

$$T_{Feder} = \Delta\theta \; k_{Feder} + T_{Inital} \quad (14)$$

wobei $\Delta\theta$ die Winkeländerung ist, k_{Feder} die Federkonstante und T_{Inital} das Initialmoment darstellt, weil die Torsionsfeder nicht im entspannten Zustand ist.
Das Drehmoment der mechanischen Anschläge lässt sich beschreiben durch die Längenänderung $\Delta\delta$ und einer stehr steifen Federkonstante $k_{mechAnschlag}$.

$$T_{mechAnschlag} = \Delta\delta \; k_{mechAnschlag} \quad (15)$$

Parallel dazu einen Dämpfer, der die Bewegungsenergie dissipiert, in dem dieser der Bewegung entgegenwirkt.

$$T_{Dämpfung} = -\omega_{rotor} \; k_{Dämpfung} \quad (16)$$

wobei das ω_{rotor} die Winkelgeschwindigkeit des Rotors darstellt und $k_{Dämpfung}$ die Dämpfungskonstante ist.
Das Reibungsmoment berechnet sich aus Winkelgeschwindigkeit und der Reibungskonstante

$$T_{Reibung} = \omega_{rotor} \; k_{Reibung} \quad (17)$$

Die Resultierende Winkelbeschleunigung des Systems lässt sich nun folgendermaßen bestimmen

$$\ddot{\theta}_{rotor} = \ddot{\omega}_{rotor} = \frac{T_{Diff}}{J} \quad (18)$$

Durch Integration der Winkelbeschleunigung kann die Winkelgeschwindigkeit bestimmt werden

$$\omega_m = \int \ddot{\theta}_{rotor} \, dt \quad (19)$$

und durch Integration der Winkelgeschwindigkeit kann der Drehwinkel bestimmt werden

$$\theta_m = \int \omega_m dt \quad (20)$$

Der Parametersatz für den BLDC-Motor ist in Tabelle 1 zusammengefasst und der Parametersatz für das mechanische Modell ist in Tabelle 2 zu sehen.

Eigenschaft	Parametername	Nominalwert [Einheit]
Spuleninduktivität	L	8.4*10^{-5} [H]
Spulenwiderstand	R	0.0625 [Ohm]
Back EMF Konstante	k_e	0.02498
Drehmomentkonstante	k_t	0.0268
Polpaarzahl	PolePairs	4

Tabelle 1: Motorparamtersatz

Eigenschaft	Parametername	Nominalwert [Einheit]
Reibungskoeffizient	FrictionCoefficient	0.01
Trägheitsmoment	Inertia	0.0001
Initialmoment der Torsionsfeder	InitialSpringTorque	0.1
Federkonstante	SpringConst	0.001
Position Anschlag links	StopperLPos	0
Position Anschlag rechts	StopperRPos	680

Tabelle 2: mechanischer Parametersatz

4. PRAKTISCHE DURCHFÜHRUNG

Die beschriebenen physikalisch mathematischen Zusammenhänge werden mit Funktionsblöcken in Matlab/Simulink nachgebildet. Die Ansteuerung des BLDC-Motors erfolgt durch die Drehrichtung als Bit und drei Hall-Sensoren. Im Datenblatt des BLDC-Motor Treiberchips ist die Kommutierungslogik beschrieben, aus der mit Hilfe der Hall-Sensoren und der Drehrichtung die Gates der Brückenschaltung angesteuert werden [2][3].

Nachdem das Modell am HIL-Prüfstand funktionsfähig ist, sind Versuche am HIL-Prüfstand möglich. Die Versuchsergebnisse werden im folgenden Kapitel mit Messungen von einem realen Prüfstand verglichen und in weiterer Folge wird das Modell auf dessen Verhalten validiert.

5. ERGEBNISSE

Bei der Validierung des Modells werden drei unterschiedliche Messungen am HIL-Prüfstand und am Prüfstand mit realem System durchgeführt. Der HIL-Prüfstand besitzt nur das Steuergerät als reale Hardware, die Leistungskomponenten, Parksperre, Kupplung und die Fahrzeugumgebung werden simuliert. Der Prüfstand mit realem System hat das komplette Aggregat aufgespannt, bestehend aus Steuergerät, elektrischer Maschine und Getriebe. Nur das Steuergerät des Fahrzeugs wird als Restbussimulation simuliert.

Im ersten Versuch wird das Stromprofil von Parksperrmodell und realer Parksperre verglichen. Der Strom soll beim Anfahren des Anschlags mit 2V im Spannungsbetrieb, denselben charakteristischen Verlauf zeigen. Abbildung 1 und 2 zeigen die Messergebnisse.

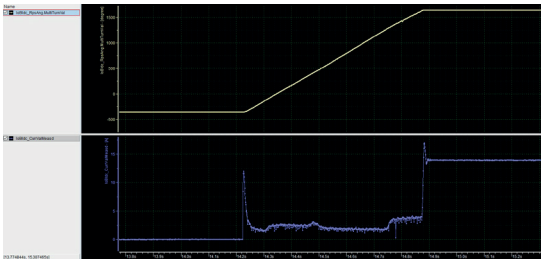


Abbildung 1: Stromprofil reale Parksperre

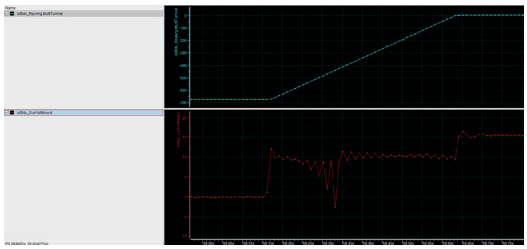


Abbildung 2: Stromprofil Parksperrmodell

Die Messungen zeigen eine hohe Ähnlichkeit beim Anlauf- und Anschlagstrom. Der Stromverlauf bei konstanter Geschwindigkeit zeigt einen großen Unterschied, dieser ist beim Modell 10 Ampere, der Verbrauch bei der realen Parksperre nur 2 bis 4 Ampere.

Der zweite Testversuch validiert das dynamische Verhalten der Parksperre und somit dessen Ein- und Auslegezeit. Die Ein- und Auslegezeit startet mit einem Befehl von der VCU und endet mit dem Erreichen des erwünschten Zustandes. In Abbildung 3 sind die Zustandssignale und die Position der Parksperre zu sehen, wobei der Parksperrbefehl in orange, der Parksperrzustand in violett und die Parksperrposition in blau eingefärbt ist. Bei dem Versuch ist zu erkennen, dass die Parksperre 0,28s zum Einlegen und 0,23s zum Auslegen benötigt.



Abbildung 3: Reaktionszeit mit realer Parksperre

Derselbe Versuch mit dem Parksperrmodell ist in Abbildung 4 zu sehen.



Abbildung 4: Reaktionszeit mit Parksperrmodell

Der Versuch zeigt, dass das Parksperrmodell nur 0,04s langsamer als die reale Parksperre einlegt und um 0,01s schneller auslegt. Somit lässt sich das Parksperrmodell kaum von der realen Parksperre unterscheiden.

Zum Abschluss wird noch eine Sicherheitsfunktion getestet, in dieser soll die ECU ein ungewolltes Auslegen verhindern. Ein ungewolltes Auslegen beschreibt eine Positionsänderung in Auslegerichtung, ohne einen Auslegebefehl zu erhalten. Das Ergebnis dieses Versuchs zeigt, dass die ECU mit realer und modellierter Parksperre korrekt reagiert und das Öffnen der Parksperre verhindert, da die Brückenschaltung deaktiviert wird [2].

6. ZUSAMMENFASSUNG

Ziel der Arbeit war es ein Parksperrmodell für den HIL-Prüfstand zu modellieren, welches die ECU nicht von einer realen Parksperre unterscheiden kann. Die Parksperre wird mathematisch nach physikalischen Gesetzen beschrieben. Anschließend wird die Parksperre in kleinere Subkomponenten unterteilt und in weiterer Folge wurden die Komponenten in Matlab/Simulink modelliert. Nachdem das Modell im HIL-Prüfstand implementiert ist, werde Versuche mit modellierter und realer Parksperre durchgeführt, um das Parksperrendmodell zu validieren. Überprüft wird dies, in dem das Stromprofil, das dynamische Verhalten und eine Sicherheitsfunktion getestet werden.

Die Ergebnisse zeigen, dass das Parksperrmodell aus funktionaler Sicht die reale Parksperre ersetzen kann. Softwareentwickler haben nun die Möglichkeit „Impact Analyse"- und Regressionstests, in denen die Parksperre Einfluss nimmt oder beeinflusst wird, schnell und kostengünstig durchzuführen. Weiters verfügt der HIL-Prüfstand nun über eine komplette Umgebungssimulation, welche das Erproben und Verbessern automatisierter Prüfläufe für den realen Prüfstand ermöglicht und folglich Kosten senkt.

7. LITERATURVERZEICHNIS

[1] Baldursson S., 2005. *BLDC Motor Modelling and Control – A Matlab Simulink Implementation*

[2] ohne Verfasser, 2017. "DRV3245Q-Q1 3 Phase Automotive Gate Driver Unit (GDU) With High-Performance Sensing, Protection and Diagnostics"

[3] ohne Verfasser, „Field-Oriented Control of PMSM Using Hall Sensor - MATLAB & Simulink Example", ohne Datum, Zugegriffen auf: https://www.mathworks.com/help/mcb/gs/foc-pmsm-using-hall-sensor-example.html



Philipp Petsch, B.Sc.

Magna Powertrain, Wienersdorfer Str. 20-24 Traiskirchen 2514, 050 444 3847

Berufliche Erfahrung
2020-2021 Magna Powertrain Business Office
Seit 2021 Magna Powertrain Systemtest
Akademische Ausbildung
2016-2019 Bachelorstudium Mechatronik Fachhochschule Wiener Neustadt
2019-2023 Masterstudium Mechatronik Fachhochschule Wiener Neustadt

FRAMEWORK FÜR DIE ANALYSE VON HOCHDIMENSIONALEN DATEN AUS DEM PBF-LB/M PROZESS

Johannes Picker

Kurzfassung: *In dieser Arbeit wurde eine Software für die Verarbeitung von Schichtdaten aus dem Laserstrahlschmelz-Prozess (PBF-LB/M) entwickelt. Durch die große Menge an Daten die entsteht ist es schwer, Zusammenhänge oder Anomalien im Prozess zu erkennen. Mittels der Software kann eine einzige gesamtheitliche Darstellung des Prozesses erzielt werden. Eine vergangene Forschung hat sich mit dem Thema beschäftigt und es wurde festgestellt, dass durch die Anwendung eines dimensionsreduzierenden Algorithmus der Prozess besser abgebildet werden kann. In dieser Arbeit wurde der Code für die Verarbeitung der Messdaten überarbeitet, um eine breite Menge von Messdaten zu untersuchen. Die Überarbeitung umfasst eine graphische Benutzeröberfläche für das Starten von Funktionen, die Auswahl der Messdaten und der Festlegung von Berechnungsparametern. Unteranderem wurde eine Methode für die Transformation der Messdaten in eine geeignete Form für die Dimensionsreduktion entwickelt und weiters auch eine Rücktransformation in Schichtbilder aus den Ergebnissen der Dimensionsreduktion ermöglicht. Diese Methodik wurde mittels eines Datensatzes, bei dem Prozessparameter gezielt verändert wurden erfolgreich validiert.*

Schlüsselwörter: *Additive Fertigung, Datenverarbeitung, Dimensionsreduktion, Laserstrahlschmelzen (PBF-LB/M), Prozessoptimierung*

1. EINLEITUNG

Die additive Fertigung spielt aufgrund der hohen Gestaltungsfreiheit und der breiten Verfügbarkeit am Markt eine große Rolle. Der Prozess verläuft meist Schicht für Schicht, bis die gewünschte Bauteilgeometrie erreicht wird. Das Fertigungsverfahren des Laserstrahlschmelzen (engl. Laser Powder Bed Fusion) (PBF-LB/M) ermöglicht die Herstellung von hochkomplexen Geometrien, die konventionell kaum möglich sind. Ausgangsmaterial beim PBF-LB/M-Verfahren ist Metallpulver. [Godec et al. 2022]

Additiv gefertigte Bauteile unterliegen aufgrund vieler Prozesseinflüsse oft einer nachträglichen Qualitätsprüfung. Insbesondere in sicherheitsrelevanten Bauteilen für Luft- und Raumfahrt ist diese Prüfung unverzichtbar. Um den Anforderungen gerecht zu werden, werden Systeme entwickelt, um die Prozesswiederholbarkeit und Reproduzierbarkeit zu verbessern. [Mahmoud et al. 2021]

Es gibt Ansätze der Prozessüberwachung, bei denen es jedoch schwer ist, klare "Regeln" für den Prozess zu definieren. Hierbei erfolgt lediglich eine Prüfung von fehlerfreien und fehlerhaften Bauteilen, jedoch ohne nähere Einsicht in den Grund für den Fehler.

2. PROBLEM- UND AUFGABENSTELLUNG

Für die Prüfung der Genauigkeit und technischen Eigenschaften nach dem Druck werden kostenintensive Verfahren wie die Computertomographie (CT) verwendet. Um dies zu umgehen werden Messsysteme verwendet, welche den Prozess im Vorfeld bewerten.

Das in dieser Arbeit untersuchte Messsystem ist mit mehreren Sensoren ausgestattet und liefert unterschiedliche Darstellungen des Prozesses. Aufgrund der hohen Datenmenge ist es menschlich kaum möglich, Zusammenhänge aus den Messdaten herauszuziehen. Um diese Daten zu vereinfachen, wurde ein Framework entwickelt, in dem Anwender:innen mithilfe einer einzigen Darstellung ein Einblick in relevante Bereiche gewährt wird. Dadurch können Auffälligkeiten hervorgehoben und leichter erkennbar gemacht werden. Eine grafische Benutzeroberfläche (GUI) in MATLAB® ermöglicht die Interaktion und umfasst das Einlesen der Messdaten, die Auswahl eines Interessensbereichs, die Dimensionsreduktion, die Ergebnisverarbeitung, die Rücktransformation in Schichtbilder und die Datenarchivierung.

Die Forschungsfragen der Arbeit betrafen die Erkennung von Prozessanomalien anhand der Daten, die Rolle der Dimensionsreduktion zur Bereitstellung von mehr Prozessinformationen und geeignete Methoden zur Rücktransformation.

3. MATERIALIEN UND METHODEN

Das Messystem liefert als Output Schichtbilder im TIFF-Format, dabei bilden die Pixel-Informationen in den Schichtbildern die aufgenommen Sensordaten ab. Die Anforderungen des Frameworks umfasst das Laden und Prozessieren der Schichtbilder. Schematisch kann das Framework in fünf Schritte unterteilt werden: Input, Pre-Processing, t-SNE, Clustering und Output.

Im Input können die Messdaten durch die Anwender:innen ausgewählt werden. Im Pre-Processing geschieht die Verarbeitung der Schichtbilder in eine geeignete Form für die weiteren Berechnungen. Für die Dimensionsreduktion wird die t-verteilte stochastische Nachbarschaftseinbettung (engl. t-distributed Stochastic Neighbor Embedding), kurz t-SNE verwendet. Der Aufbau einer t-SNE Eingangs-Matrix gestaltet sich als D×M Matrix, wobei D die Anzahl der Datenpunkte und M die Anzahl der Merkmale umfasst. Die Schichtbilder wurden so transformiert, dass die X-Y-Z Messpunkte als Datenpunkte fungieren und die gemessenen Sensordaten als Merkmale festgehalten werden. Durch diese Struktur ist die Durchführung der Dimensionsreduktion mit dem t-SNE Algorithmus möglich. Das Ergebnis umfasst ein Streudiagramm, in dem Ähnlichkeiten aus dem hochdimensionalen Raum als Punktegruppierungen gesehen werden können. Um aus der Streudiagramm-Darstellung auf Schichtbilder zu kommen, wurde der DBSCAN-Clustering Algorithmus angewendet, um Punktegruppierungen zu identifizieren. Die Identifizierung von den Gruppierungen hilft in der Diskretisierung von Bereichen in den Schichtbildern. Nachdem die Cluster gefunden wurden, werden diese verwendet, um die Schichtbilder einzufärben. Als Output liefert das Framework wiederum Schichtbilder im TIFF-Format.

4. PRAKTISCHE DURCHFÜHRUNG

Es wurde eine Software entwickelt, welche den gestellten Anforderungen entspricht. Es wurde ein GUI für Input, Pre-Processing, t-SNE, Clustering und Output entwickelt. Weiters wurden ein Interface für das Festlegen von Berechnungsparametern und das Visualisieren von Ergebnissen erreicht. Bei der Visualisierung der Ergebnisse wurde weiters eine 3D-Darstellung der Schichtbilder programmiert.

Für das Testen der Software wurde ein Druck-Datensatz untersucht, der aus dem FFG-Projekt „LBMCheck“ entstammt. Als Testkörper wurden quaderförmige Bauteile, welche in sechs Höhen-Bereiche unterteilt wurden, gedruckt. Pro Bereich wurden Prozessparameter, wie beispielsweise die Laserleistung des Druckers variiert. Eine Validierung des Frameworks geschieht, wenn diese Bereiche gesondert erkannt werden.

Für die Validierung wurden vier Darstellungen, die mit dem Messsystem erstellt wurden, mithilfe des Frameworks untersucht. Es wurden drei Höhen-Bereiche zu je 20 Schichten prozessiert. Um optimale Parameter für die Berechnung zu finden, wurde weiters eine Parameteroptimierung durchgeführt. Es wurde ein Durchlauf des Frameworks durchgeführt und die Schichtbilder exportiert.

5. ERGEBNISSE

Die Ergebnisse des Frameworks sind in Abb. 1 und Abb. 2 ersichtlich. In Abb. 1 ist das GUI für das Festlegen der Parameter der DBSCAN Funktion und den Export der Schichtbilder demonstriert. Links in der Abbildung ist das Ergebnis der t-SNE dargestellt, wobei die Erkennung der Bereiche durch die Clusteranalyse mit DBSCAN geschieht. Anhand dieser Bereiche werden die resultierenden Schichtbilder eingefärbt. In Abb. 2 ist das GUI für die Analyse der exportierten Bilder ersichtlich, auf der linken Seite ist ein Histogramm über die Cluster im Bau-Bereich dargestellt. Mit den darunterliegenden Messschiebern können Cluster weiters eingegrenzt werden, um Bereiche einzeln zu untersuchen. Rechts in der Abbildung ist die 3D-Visualiserung des Testkörpers dargestellt. Aus der Abbildung ist erkennbar, dass die drei Höhen-Bereiche klar voneinander getrennt sind, wodurch die Funktionalität des Frameworks validiert ist. Weiters sind auch Bereiche innerhalb der Schichten hervorgehoben.

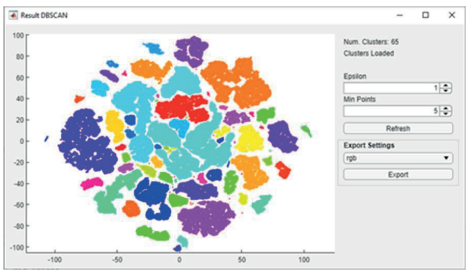


Abb. 1: GUI und Ergebnis der Clusteranalyse mit DBSCAN

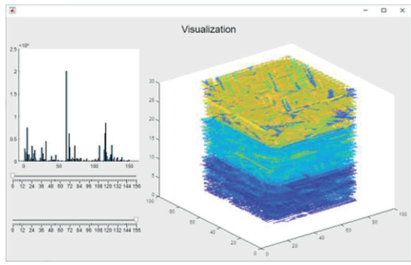


Abb. 2: GUI für die 3D-Visualisierung der Ergebnisse

6. ZUSAMMENFASSUNG UND AUSBLICK

Die Masterarbeit hatte das Hauptziel, ein anwendbares Framework zur Analyse von Prozessdaten zu entwickeln. Die Gesamtziele wurden erreicht, darunter die Verbesserung der Prozessvisualisierung, die Erkennung von Prozessvariationen sowie die Schaffung eines nutzerfreundlichen Interfaces. Die Datenverarbeitung erfolgte über eine Pipeline mit t-SNE-Dimensionsreduktion und DBSCAN-Clustering. Die Rücktransformation der Daten aus der t-SNE-Darstellung wurde erfolgreich umgesetzt. Das Framework ermöglichte eine Abgrenzung von Prozessbereichen und gewährte tiefere Einblicke.

Das entwickelte Framework ermöglicht die Darstellung von Schichtbildern aus mehreren Kennwerten einer PBF-LB/M-Maschine, aufgenommen durch ein schichtbasiertes Messsystem. Es bietet Anwendungspotenzial zur Anomalie-Erkennung im Druckprozess und fördert das Verständnis des Prozesses durch die Vereinfachung großer Datenmengen und Identifikation von Mustern. Eine mögliche Erweiterung des Frameworks beinhaltet ein GUI zur Kombination von t-SNE-Ergebnissen mit 3D-Visualisierung, wodurch tiefere Einblicke in die Datenanalyse ermöglicht werden könnten. Es bildet eine Grundlage für die Visualisierung und Optimierung, mit Potenzial für Erweiterungen und Verbesserungen in zukünftiger Forschung.

Um anderen Anwendern:innen die Analyse zu ermöglichen ist der gesamte Quellcode der Software auf GitHub veröffentlicht unter:
https://github.com/JP4799/process_insight_framework

7. LITERATURVERZEICHNIS

Godec, Damir, Gonzalez-Gutierrez, Joamin, Nordin, Axel, Pei, Eujin und Ureña Alcázar, Julia(hrsg) (2022): *A Guide to Additive Manufacturing*, Cham, Springer International Publishing

Mahmoud, Dalia, Magolon, Marcin, Boer, Jan, Elbestawi, M. A. und Mohammadi, Mohammad Ghayoomi (2021): *Applications of Machine Learning in Process Monitoring and Controls of L-PBF Additive Manufacturing: A Review* In: Applied Sciences, Band 11, Ausgabe 24, DOI: 10.3390/app112411910.



Johannes Picker, BSc.

FH Wiener Neustadt, Johannes-Gutenberg-Straße 3, johannes.picker@outlook.com

2021-2023: Masterstudium Mechatronik, Spezialisierung „Additive Manufacturing“
2018-2021: Bachelorstudium High-Tech-Manufacturing, FH Campus Wien

DESIGN OF A TUNABLE MICROWAVE SOURCE FOR INTEGRATED QUANTUM TECHNOLOGY

Felix Schinkovits

Abstract: *Voltage-controlled oscillators (VCOs) serve as fundamental building blocks in many integrated circuits and are used in various applications, such as telecommunications, digital, measurement, and sensor technology. This thesis delves into the analysis and design of a low-phase-noise VCO for integrated quantum technology, such as atomic clocks and applications in quantum information processing. The desired fundamental frequency of 6.835 GHz corresponds to the transition frequency of the highly precise hyperfine transition of the ground state of Rubidium. To achieve this frequency, an LC topology with an integrated inductor, whose characteristics were numerically simulated, was selected. By employing a voltage-controlled varactor circuit, the output frequency can be adjusted within a range of +/- 30 MHz around the resonant frequency. For the physical design, a 180-nm CMOS process with a supply voltage of 3.3 V was chosen. Simulated results demonstrate a phase noise of approximately -115 dBc/Hz at an offset frequency of 1 MHz.*

Keywords: *LC-VCO, CMOS, on-chip inductance, low phase noise*

1. INTRODUCTION

The versatile landscape of quantum mechanics unfolds an array of multifaceted applications, from foundational science to state-of-the-art technologies [1]. This work contributes to a bigger project that focuses on the development of components for quantum technological devices using ultra-cold atoms. Such devices can be subsumed under the term “atomtronics” and are used in applications such as quantum computing, quantum simulation, and quantum sensing [2]. The aim of this work is the development of an oscillator with a resonating frequency of *6.835 GHz* that can be used in quantum computing (which is the main application of the oscillator and the application for which the oscillator was optimized for) and quantum sensing applications, such as atom clocks.

In the last two decades, quantum computing has become an emergent field of research. Although systems with practical use have not yet been developed, experimental setups show promising results. It is evident that large-scale quantum computers can solve certain tasks exponentially faster than any “traditional” computer [3]. A quantum computer uses quantum mechanics, especially the principles of superposition, coherence, and entanglement to manipulate states and solve computational problems. The basic information unit in a quantum computing system is the qubit (quantum bit). Similar to a classical bit, the qubit is a two-state system, with states 0 and 1. While a classical bit has to be in one specific state, the qubit, due to its quantum mechanical nature, can be in a coherent superposition of both states simultaneously [4]. According to the physicist David P. DiVincenzo [5], five conditions must be met in order to construct qubits for quantum computation:

1. scalability: the possibility to build a large number of well-characterized qubits
2. initialization: the ability to initialize the state of the qubits to a well-defined quantum state
3. long decoherence time: the time in which quantum-mechanic behavior occurs must be long compared to the gate time
4. universal set of quantum gates: to allow for all possible quantum algorithms
5. measurement procedure: a qubit-specific measurement capability to get the result of a calculation

The project aims to efficiently fulfil the second condition (ability to initialize the qubits) of DiVincenzo’s criteria. The qubits in this project are realized by using the hyperfine states F=1 and F=2 of the $5^2S_{1/2}$ ground state of Rubidium (Rb) (see the energy level scheme in Figure 1), where F=1 represents the |0> (ground state) and F=2 represents the |1> (excited state). To precisely control and manipulate the states of the atoms, their thermal movement must be reduced by cooling the atoms down. This project uses a technique known as “magneto-optical trapping”, which uses a combination of a laser and magnetic trapping in order to cool the atoms [6]. First, a laser with a distinct frequency is used. When photons from the laser are absorbed by the atom, the atoms gain a momentum opposite to the laser beam’s propagation direction, causing the atoms to slow down and hence cooling them. Second, a magnetic gradient field is used to split up the hyperfine structure levels (F-levels) into Zeeman sublevels (mf-levels). The F=1 level is split into mf=-1, 0, 1 and the F=2 level is split into mf=-2, -1, 0, 1, 2. When the laser cooled atoms enter regions in which the magnetic field is at a minimum, the magnetic force stops them from escaping this region and traps them in place once they are brought into the trappable states. The desired trapped states are F=1, mf=-1 and F=2, mf=1. Atoms that are too hot consequently are filtered out by using evaporative cooling, thereby removing the hottest atoms from the trap so that the coldest part of the ensemble remains [7].

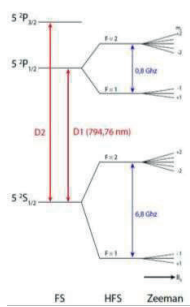


Fig. 1: Energy level scheme of ^{87}Rb with the fine structure (FS), the hyperfine structure (HFS) and the Zeeman levels [8]

When the atoms are “cold” enough then the states F=1 and F=2 can be addressed separately and atoms in the ground state can be transferred to the excited state by using a microwave field with a frequency equal to the hyperfine frequency of $v_0 \approx 6.835$ GHz [6]. Next to the microwave field, a radio-frequency field that is phase-locked to the microwave field must be provided since transition should only occur between trapped states. A transition from F=1 mf=-1 to F=2, mf=1 requires a change in the Zeeman level “mf” of two. The microwave field can only change the Zeeman level by one, so one additional phase-locked photon is required to change the Zeeman level by two in total. This process is commonly known as “two-photon” transition [6].

In order to measure if the atoms are in the excited or ground state a narrowband laser probe is used. If the microwave frequency is exactly at v_0, then a maximal number of atoms is converted into the excited state and the absorption of light is maximal. If the frequencies do not match the absorption level is lower. A photodetector transforms the transmitted light into a proportional voltage, which is then used in a feedback loop to alter the frequency of the microwave source in order to match v_0 [9].

2. PROBLEM STATEMENT

This work focuses on the development of a low-phase noise microwave source using Complementary Metal Oxide Semiconductor (CMOS) technology, that is capable of generating an output signal with a frequency that is equal to the hyperfine frequency v_0 of approximately 6.835 GHz. This microwave source should be based on a voltage-controlled oscillator (VCO) whose ground frequency f0 should be 6.835 GHz that should be tunable by +/- 25 MHz by applying a control voltage.

3. METHODS

Over the years, numerous different oscillator topologies, have been introduced. This thesis focuses on LC-oscillators since they are commonly used in the required frequency range and offer a low phase noise profile that is necessary for an application within phase-locked loops.

An oscillator can be viewed as two separate systems (see Figure 2): a lossy resonator and an active circuit that cancels the loss [10]. Without the active circuit, the oscillation would not be sustained, because some energy in the tank is transformed into heat with each cycle. If an active circuit provides the same amount of energy to the LC-tank that is “burned” by the resistance in the tank, then the signal would swing with full amplitude indefinitely. In fact, the active circuit must cancel out the effect of R_p in Figure 2 by having an input resistance of −Rp or less [10].

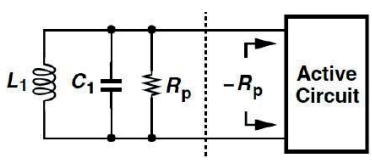


Fig. 2: LC-tank with active circuit that cancels the loss in the tank [10]

This active circuit can be accomplished by cascading two common-source stages in a loop, as can be seen in Figure 3.

When examining the closed loop, we see that each of the two stages contributes with a phase shift of 180°, summing to 360°, which results in a positive feedback system. To keep up the oscillation, a loop gain equal to or greater than unity is necessary. This is given if the transconductance of the MOSFETs M_1 and M_2 multiplied with the resistance of the tank, which equals the loop gain, is greater than unity ($g_m \cdot R_p \geq 1$) [10].

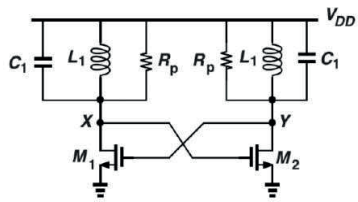


Fig. 3: Circuit of cross-coupled LC-oscillator [10]

In order to improve the characteristics of the VCO, a pair of PMOS transistors (M_3 and M_4 in Figure 4) can be introduced next to the pair of NMOS transistors (M_1 and M_2 in Figure 4). This topology offers several advantages compared to the circuit with only one transistor pair:

- greater voltage swing and hence less time during noise from the transistors is injected into the tank
- symmetric current swings within the tank
- symmetric output can be achieved
- faster switching and better rise/fall time symmetry [11].

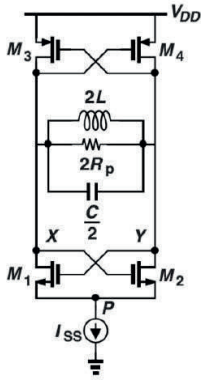


Fig. 4: Oscillator with complementary cross-coupled pairs [10]

4. DESIGN OF THE VCO

The design of the VCO, was divided into three steps. First, a customized inductance for the LC-tank was designed. Second, the other circuit parts, namely the cross-coupled pairs, the current source, and the varactor were designed. Lastly, the layout was created, and final simulations were conducted.

4.1 INDUCTANCE DESIGN

For the LC-tank a customized inductance was designed. Therefore, the inductance value was first approximated with a heuristic formula from Wheeler [12]. The different electrical parameters of the inductance then were determined in the

multiphysics tool COMSOL[1]. The final geometry of the inductance can be seen in Figure 5. It consists of an octagonal two-turn spiral with a wire width of 5 µm, a spacing of 5 µm and an outer diameter of 130 µm, an isolation layer out of SiO_2, and a patterned ground shield.

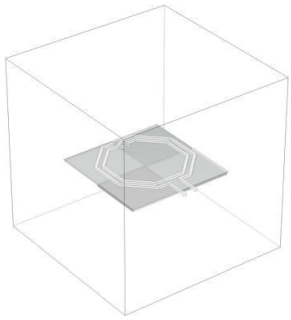


Fig. 5: Geometry of inductance in COMSOL

COMSOL was used to numerically determine the inductance value, the series resistance, the capacitance between the windings and the capacitance between the inductor and the groundshield at a frequency of *6.835 GHz*. The schematic model that was used for the following spice simulations can be seen in Figure 6.

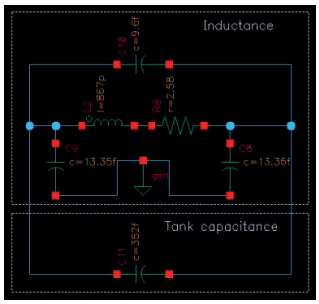


Fig. 6: Schematic model of the LC-tank

4.2 DESIGN OF THE CROSS-COUPLED PAIRS

As described in chapter 3, an active circuit with an input resistance that is equal to the resistance in the tank or less is required for steady oscillation. This negative resistance is provided by two cross-coupled pairs (see Figure 4). In order to determine the minimal required size of the MOSFETs to ensure oscillation, a transient analysis in combination with a parameter sweep of the transistor widths was conducted. Since the charge carriers in PMOS-devices are holes and the mobility of holes in silicon is typically approximately 2.7 times lower than that of electrons (which are the charge carriers in NMOS-devices), the width of the PMOS-cross coupled pair will be designed 2.7 times larger than that of the NMOS-cross coupled pair to allow for comparative rise and fall times [13].

In Figure 7 the results of the simulation are displayed. One can see the output voltage for three different widths of the MOSFETs (the size of the PMOS-MOSFETs was chosen as 2.7 times the size of the NMOS-devices, as described above). The waveforms look filled, because the oscillation frequency is in the GHz range, but a simulation time of 150 ns is displayed. One can clearly see that with a width of 6 µm of the NMOS-MOSFETs the negative resistance is too low and the oscillation decays after a few nanoseconds. With a width of 6.5 µm the oscillation builds up until it saturates. This shows that the negative resistance of the cross-coupled pairs exceeds the losses in the tank. Although oscillation occurs, the time until a steady state is reached equals approximately 100 ns, which is quite long. With a width of 7 µm the oscillation starts-up fast and after approximately 35 ns a steady state is reached. Hence, the minimum size for reliable oscillation is approximately 7 µm.

The cross-coupled pairs not only provide the necessary negative resistance, but also are a significant source of phase noise. On the one hand, wider transistors reduce the flicker noise, which would be a reason to use wide transistors. On the other hand, the transistor width is proportional to the thermal noise in MOSFETs. Hence, a compromise between the flicker noise and the thermal noise had to be made. Shin et.al[14] determined the optimal width of the NMOS cross-coupled pair, as W/L = 30 µm/0.18 µm for a 3.5 mA bias current in a 0.18µm-process. Unfortunately, the minimal gate length for MOSFETs that function with 3.3 V in the used manufacturing process is 0.35 µm for NMOS-devices and 0.30 µm for PMOS-devices. Keeping the W/L ratio the same as in Shin et.al. [33] would result in a significant layout footprint of the cross-coupled pairs. In order to keep the layout footprint of the cross-coupled pairs rather small, the NMOS-MOSFETs were designed with a W/L ratio of 30 µm/0.35 µm and the PMOS-MOSFETs were designed with a W/L ratio of 81 µm/0.35 µm.

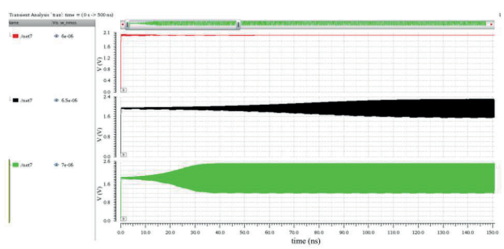


Fig. 7: Output voltage over width of transistors

4.3 DESIGN OF THE CURRENT SOURCE

In order to ensure robust operation, proper biasing of the cross-coupled pairs is necessary. This biasing was achieved by introducing a current source, that establishes a fixed current. The current source is a big contributor to the phase noise of the VCO. That is why the two topologies (PMOS vs. NMOS current sources) were compared based on the phase noise characteristic. Using the same parameters for both the VCO with the NMOS- and the PMOS current source, the phase noise characteristic was analyzed. It showed that for low offset frequencies the VCO that uses the PMOS current mirror exhibits significantly lower phase noise than the VCO with the NMOS current mirror. The difference at an offset frequency of 10 kHz is approximately 10 dBc/Hz. This difference can be explained by the lower flicker noise in PMOS devices, which is dominant at low offset frequencies. The specifications of the devices that were used for the current source can be seen in Figure 8. A current ratio of ten was achieved by designing the gate width of transistor M37 ten times wider than that of transistor M38. The resistor R1 sets the biasing current to 3.5 mA.

4.3 DESIGN OF THE VARACTORS

The frequency of the oscillator can be changed by changing the total capacitance in the circuit. This change in

[1] www.comsol.com

capacitance can be realized by using a voltage-dependent capacitor, called "Varactor". A varactor can be realized by using a normal MOSFET, connecting the drain and source terminal, and using the voltage-dependent gate capacitance. The gate capacitance varies with the voltage between the gate and the source (VGS) [15]. Unfortunately, this behavior is non-monotonic. As a result, the output frequency of the VCO would first rise and then fall with rising VGS. This behavior is especially unwanted when using a VCO in a phase-locked loop since this nonmonotonic behavior cannot be easily implemented in the electronic that alters the control voltage. The above-mentioned issue can be avoided by placing an NMOS-transistor in an n-well. This structure is named "accumulation-mode MOS varactor" (AMOS-varactor) [16].

Unfortunately, the AMOS-varactors for which a Cadence cell is provided in the used process (XH018 from XFAB), require a minimal gate length of around 6 µm for stable operation. Therefore, a varactor with a gate width of 50 µm, a gate-length of 6 µm, and five gate fingers was used. This varactor introduced enough capacity change to achieve a tuning range of +/- 30 MHz around a ground frequency of 6.835 GHz.

5. RESULTS

The final schematic of the complementary cross-coupled VCO, that was used for the final simulations, can be seen in Figure 8. It consists of six distinct parts. The current source provides a constant bias current in order to properly bias the transistors. The PMOS and NMOS CC-pairs provide the negative resistance that is necessary to cancel out the losses in the tank and hence to maintain oscillation. The on-chip inductance was modeled by an inductivity in series with a resistor, a parallel winding capacitance, and the capacitances between the inductance and the ground shield. These components were approximated by using heuristic formulas and the multiphysics tool COMSOL. The oscillation frequency of the VCO can be tuned by adjusting the control voltage at the varactor.

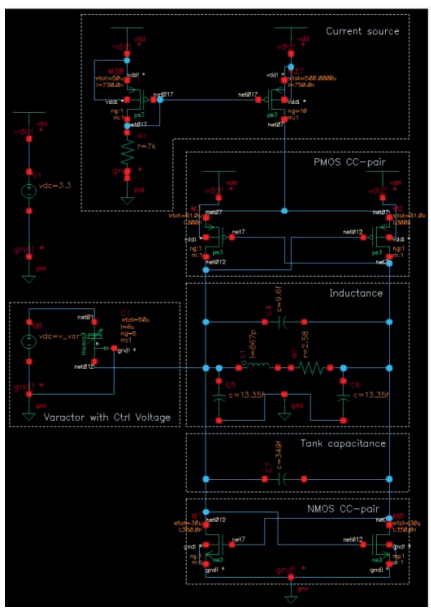


Fig. 8: Final circuit including parasitic components

The final output voltage can be seen in Figure 9. It shows that the oscillation reaches its maximal amplitude of approximately 2 V at around 4 ns. Since for the planned application the time until the frequency changes occur at the output when the control voltage is changed is not critical, this circuit is suitable for the application at hand regarding that aspect.

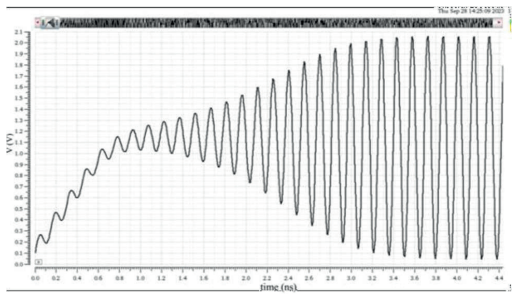


Fig. 9: Output voltage during startup

The VCO can be tuned by changing the control voltage at the varactor. In order to achieve linear behavior, the circuit was tuned to operate in the varactor region in which the C/V behavior is nearly linear. The resulting frequency over control-voltage plot can be seen in Figure 10. By changing the control voltage from around approximately 0.21 V to approximately 1.03 V the frequency of the VCO can be changed from *6.865 GHz* to *6.805* GHz. When approximating the relationship between the frequency and a control voltage with a linear regression this results in a K_{VCO} of *73.13 MHz/V.*

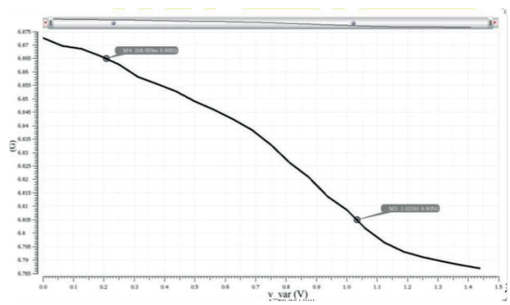


Fig. 10: Tuning range of frequency

The phase noise characteristic of a VCO is a crucial factor that describes the quality of the oscillation. It is a measure for how much power within a one Hertz range the signal carries at frequencies other than the desired oscillation frequency. The phase noise characteristic of the VCO can be seen in Figure 11. The phase noise profile falls with a slope of around −30 dBc/Hz per decade until an offset frequency of around 100 kHz. For higher frequencies, the phase noise profile flattens and shows a slope of around −20 dBc/Hz per decade. For even higher frequencies than displayed in Figure 11 the phase noise profile would flatten out until only white noise occurs. This phase noise profile is typical for LC-VCOs, as it results out of the different noise sources within the circuit. The designed VCO shows a phase noise of approximately −61 dBc/Hz at an offset frequency of 10 kHz, approximately -91 dBc/Hz at an offset frequency of 100 kHz and approximately −115 dBc/Hz at an offset frequency of 1 MHz.

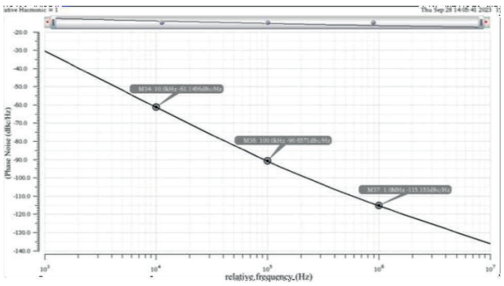


Fig. 11: Final phase noise profile

The layout of the designed VCO can be seen in Figure 12. The footprint of the design is 169 µm wide and 220.5 µm high. The largest part of the footprint is occupied by the on-chip inductance. Beside the tank and the groundshield, the layout consists of four PMOS-transistors, two NMOS transistors, one NMOS-varactor, and one resistor. In order to reduce the wiring, all NMOS devices were placed in a P-guardring that connects the MET1 to the p-substrate. By connecting the guardring to ground, the Bulk terminals of all NMOS-devices within the ring are connected to ground. The same principle was used to connect the Bulk terminals of all PMOS-devices. Instead of a P-guardring a N-guardring was used that connects the MET1 layer to an N-well. To ensure that all PMOS-devices are placed within an N-well, an N-well area had to be drawn within the ring.

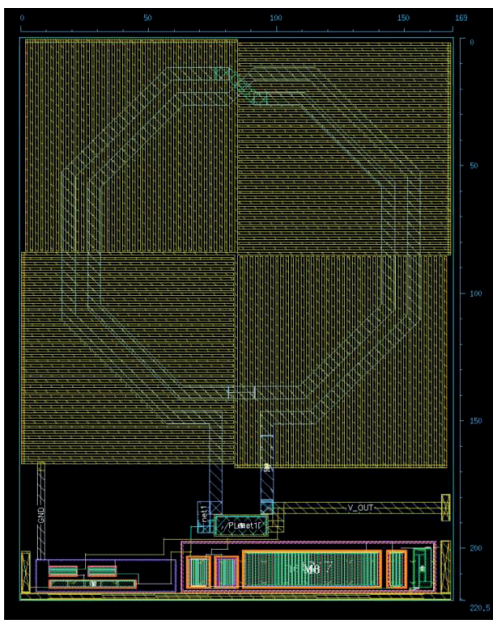


Fig. 12: Layout of complete VCO

6. SUMMARY AND FUTURE OUTLOOK

VCOs are typically compared based on the phase noise at a certain offset frequency. VCOs that are described in literature mostly report the phase noise at an offset frequency of 1 MHz. In table 1, the process, the operating frequency, the phase noise at an offset frequency of 1 MHz, and whether the phase noise was simulated or measured for different VCOs that were found in publications are displayed. It shows that the phase noise of the designed VCO is comparable with the phase noise of the VCOs described in literature. One possible reason for why a lower phase noise profile could not be reached is the process depended higher minimal gate width for 3 V devices. The transistor geometries highly influence the phase noise (especially the 1/f noise) of the transistor.

CMOS-Process	Frequency (GHz)	Phase noise @1MHz (dBc/Hz)	Sim./Meas.	Reference
180 nm	5.32	-116.00	Meas.	[17]
180 nm	5.20	-113.70	Meas.	[18]
180 nm	5.00	-117.70	Sim.	[19]
180 nm	5.80	-117.00	Meas.	[20]
180 nm	7.00	-103.20	Sim.	[21]
180 nm	6.80	-115.15	Sim.	This work

Tab. 1: Comparison of phase noise results to previously reported VCOs from literature

This work culminates in the design of a VCO that fulfills the requirements for the application in quantum technological devices, such as quantum computing, quantum simulation, and quantum sensing, using ultra-cold atoms. Although the most important parasitic components were determined, it is crucial to carry out a post-layout simulation within Cadence in order to ensure that the VCO works within the required frequency range. In the whole development process, it became evident that already small parasitic capacitances (within few femtofarad) could significantly move the frequency range. After the post-layout simulation, the integration of the circuit into the overlying system and the manufacturing and testing of the chip would be the next steps.

Further research regarding the optimization of the VCO could build on the following topics: First, the inductance could be replaced by a more sophisticated inductance (such as stacked inductances over more layers as described in[22]) that has a smaller footprint. This could yield a higher quality factor and hence better phase noise performance. Second, the biasing current source could be improved by using a more complex topology, such as in [23]. Third, the varactor model used in this VCO requires a minimum gate length of 6 µm for stable operation within the whole required tuning range. This again negatively affects the phase noise behavior. Further optimization could build on a varactor that does not require a minimum gate length or on a more complex varactor typology, such as the one described in [24]. Moreover, the phase noise behavior could be improved by using filtering techniques at the current source and the varactor input as described in [14].

7. REFERENCES

[1] S. K. Sood and Pooja, “Quantum Computing Review: A Decade of Research,” *IEEE Trans Eng Manag*, pp. 1–15, 2023, doi: 10.1109/TEM.2023.3284689.

[2] L. Amico *et al.*, “Roadmap on Atomtronics: State of the art and perspective,” *AVS Quantum Science*, vol. 3, no. 3, Sep. 2021, doi: 10.1116/5.0026178.

[3] G. Arun and V. Mishra, “A review on quantum computing and communication,” in *2014 2nd International Conference on Emerging Technology Trends in Electronics, Communication and Networking*, IEEE, Dec. 2014, pp. 1–5. doi: 10.1109/ET2ECN.2014.7044953.

[4] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information*. Cambridge University Press, 2012. doi: 10.1017/CBO9780511976667.

[5] D. P. DiVincenzo, "The Physical Implementation of Quantum Computation," *Fortschritte der Physik*, vol. 48, no. 9–11, pp. 771–783, Sep. 2000, doi: 10.1002/1521-3978(200009)48:9/11<771::AID-PROP771>3.0.CO;2-E.

[6] J. Fortágh and C. Zimmermann, "Magnetic microtraps for ultracold atoms," *Rev Mod Phys*, vol. 79, no. 1, pp. 235–289, Feb. 2007, doi: 10.1103/RevModPhys.79.235.

[7] R. Shankar, *Principles of Quantum Mechanics*. New York, NY: Springer US, 1994. doi: 10.1007/978-1-4757-0576-8.

[8] F. Borgosano, "Studies of the influence of thermodynamical parameters on the production rate of hyperpolarised 129Xe and the degree of hyperpolarisation," Sep. 2012. [Online]. Available: https : / / archiv . ub . uni-heidelberg.de/volltextserver/13703/ (visited on Oct. 15, 2023).

[9] R. Elvin *et al.*, "Towards a compact, optically interrogated, cold-atom microwave clock," *Advanced Optical Technologies*, vol. 9, no. 5, pp. 297–303, 2020, doi: 10.1515/aot-2020-0022.

[10] B. Razavi, *Design of CMOS Phase-Locked Loops*. Cambridge University Press, 2020. doi: 10.1017/9781108626200.

[11] J. Hu Ying, "A comparative study of low phase noise VCOs," 2006. Accessed: Jul. 10, 2023. [Online]. Available: https://core.ac.uk/download/pdf/211513145.pdf (visited on Oct. 15, 2023).

[12] H. A. Wheeler, "Simple Inductance Formulas for Radio Coils," *Proceedings of the IRE*, vol. 16, no. 10, pp. 1398–1400, Oct. 1928, doi: 10.1109/JRPROC.1928.221309.

[13] E. Bruun, *CMOS Analog IC Design: Fundamentals*, 2nd ed., vol. 618, no. November. Book Boon, 2019.

[14] Y. Shin, T. Kim, S. Kim, S. Jang, and B. Kim, "A Low Phase Noise Fully Integrated CMOS LC VCO Using a Large Gate Length pMOS Current Source and Bias Filtering Technique for 5-GHz WLAN," in *2007 International Symposium on Signals, Systems and Electronics*, IEEE, Jul. 2007, pp. 521–524. doi: 10.1109/ISSSE.2007.4294527.

[15] B. Razavi, *RF Microelectronics (2nd Edition) (Prentice Hall Communications Engineering and Emerging Technologies Series)*, 2nd ed. USA: Prentice Hall Press, 2011.

[16] B. Razavi, "A 1.8 GHz CMOS voltage-controlled oscillator," in *1997 IEEE International Solids-State Circuits Conference. Digest of Technical Papers*, IEEE, pp. 388–389. doi: 10.1109/ISSCC.1997.585453.

[17] Y. Shin, T. Kim, S. Kim, S. Jang, and B. Kim, "A Low Phase Noise Fully Integrated CMOS LC VCO Using a Large Gate Length pMOS Current Source and Bias Filtering Technique for 5-GHz WLAN," in *2007 International Symposium on Signals, Systems and Electronics*, IEEE, Jul. 2007, pp. 521–524. doi: 10.1109/ISSSE.2007.4294527.

[18] Young-Jin Moon, Yong-Seong Roh, Chan-Young Jeong, and Changsik Yoo, "A 4.39–5.26 GHz LC-Tank CMOS Voltage-Controlled Oscillator With Small VCO-Gain Variation," *IEEE Microwave and Wireless Components Letters*, vol. 19, no. 8, pp. 524–526, Aug. 2009, doi: 10.1109/LMWC.2009.2024846.

[19] W. Ying, P. Qin, J. Jin, and T. Mo, "A 1mW 5GHz current reuse CMOS VCO with low phase noise and balanced differential outputs," in *2011 International Symposium on Integrated Circuits*, IEEE, Dec. 2011, pp. 543–546. doi: 10.1109/ISICir.2011.6132017.

[20] C. Guo, J. Hu, S. Zhu, H. Sun, and X. Lv, "A 5-GHz low-phase-noise CMOS LC-VCO for China ETC applications," in *2011 IEEE International Conference on Microwave Technology & Computational Electromagnetics*, IEEE, May 2011, pp. 267–269. doi: 10.1109/ICMTCE.2011.5915508.

[21] S. S. S and S. S. Yellampalli, "Design of 6.7 GHz ˜ 7. 518 GHz Cross Coupled LC-VCO in 180nm CMOS technology," in *2021 5th International Conference on Computing Methodologies and Communication (ICCMC)*, IEEE, Apr. 2021, pp. 665–671. doi: 10.1109/ICCMC51019.2021.9418027.

[22] R. Murakami, K. Okada, and A. Matsuzawa, "A 484-µm2 21-GHz LC-VCO beneath a stacked-spiral inductor," in *The 3rd European Wireless Technology Conference*, 2010, pp. 245–248.

[23] T. Azadmousavi, E. N. Aghdam, and J. Frounchi, "A Low Power Current-Reuse LC-VCO with Self Body-Bias Schema," in *Electrical Engineering (ICEE), Iranian Conference on*, IEEE, May 2018, pp. 294–299. doi: 10.1109/ICEE.2018.8472561.

[24] M. Fang and T. Yoshimasu, "A −197.3-dBc/Hz FoM $_T$ Wideband LC-VCO IC With a Single Voltage-Controlled IMOS-Based Novel Varactor in 40-nm CMOS SOI," *IEEE Trans Microw Theory Tech*, vol. 68, no. 10, pp. 4116–4121, Oct. 2020, doi: 10.1109/TMTT.2020.2990441.



Felix Schinkovits, B.Sc.

Attended the Bachelor's and Master's Program in Mechatronics at the University of Applied Sciences Wiener Neustadt and the Bachelor's Program in Business Administration at the Vienna University of Economics and Business; Completed various internships in hardware design, software design, and consulting.

MODELLIERUNG DER ENTSTEHUNG UND ABTRAGUNG VON TRIBOSCHICHTEN IN REIBPROZESSEN

Seidl, V.

Kurzfassung: *Das Ziel dieser Arbeit ist die Erstellung eines realistischen Modells zur Entstehung und Entfernung von Triboschichten, das auf der Grundlage gemessener Werte angepasst und interpretiert wird. Dies erfolgt auf Basis von Ratengleichungen, die geometrischen Überlegungen entspringen, welche durch die Beobachtung des Schichtbildungsprozesses entstanden sind. Im Weiteren ist es erforderlich den Ratenkonstanten eine physikalische Bedeutung zuzuweisen, wobei hier die Temperaturabhängigkeit im Fokus steht. Die zu Grunde liegenden Forschungsfragen orientieren sich hierbei an der Realisierbarkeit eines solchen Modells bzw. dessen Einschränkungen und Vor- und Nachteile. Im Rahmen des Identifikationsprozesses wurde das Levenberg-Marquard-Verfahrens, der Downhill-Simplex-Algorithmus und die Partikelschwarmoptimierung, implementiert und im Rahmen der ersten Versuche verglichen. Hierbei konnte gezeigt werden, dass diese nahezu gleich gute Ergebnisse liefern, wobei für die weitere Parameteridentifikation die Partikelschwarm-optimierung verwendet wird. Des Weiteren konnte gezeigt werden, dass die verwendet Datensätze erfolgreich durch das vorgestellte Modell dargestellt werden können. Im Weiteren konnte die Temperaturabhängigkeit der Wachstumskonstante modelliert werden.*

Schlüsselwörter: *Wachstum und Abtrag von Triboschichten, Ratengleichungen, Parameteridentifikation*

1. EINLEITUNG

Bei Triboschichten handelt es sich also um eine durch komplexe physikalische und chemische Vorgänge gewachsene Schicht, die im Bereich der Grenzschmierung in Reibkontakten entstehen können. Dieses System ist durch eine große Anzahl an Metall-Metall-Kontakten gekennzeichnet, zwischen denen sich lediglich einige Schmiermitteldepots befinden [1]. Hierbei wird die aufgebrachte Last nur über einen Bruchteil der Kontaktfläche übertragen, wodurch extreme Drücke an den Berührungspunkten auftreten. Auch wenn das Schmiermittel keinen hydrodynamischen Auftrieb und keine Last-unterstützung bieten kann, fungiert es als Trägersubstanz von oberflächenaktiven Komponenten, den Additiven. Diese reagieren an den Oberflächen der Basissubstrate und bilden einen vor Verschleiß schützenden Tribofilm [2]. Vormals glaubte man, dass Tribofilmwachstum nur bei Festkörper-Festkörper Berührung möglich ist. Allerdings konnte gezeigt werden dass, Tribofilme selbst in hydrodynamischen Systemen Filme ausbilden, wenn die Scherspannung groß genug ist [3]. Weitere wichtige Einflussfaktoren auf die Filmbilddung sind die Temperatur, die Zusammensetzung des Öls und die Beschaffenheit der Oberfläche und des Bulkmaterials.

2. METHODEN

In dieser Arbeit wurden drei Methoden zur Parameteridentifikation beschrieben und in Matlab umgesetzt. Bei diesen Methoden handelt es sich um das Levenberg-Marquardt Verfahren, den Downhill-Simplex-Algorithmus und die Partikelschwarmoptimierung. Diese wurden in einem Einführenden Beispiel gegenübergestellt, woraus sich ableiten lässt, dass alle erwähnten Methoden valide und ähnlich gute Ergebnisse liefern. Diese Arbeit behandelt hierzu ausschließlich die Frage nach einer geeigneten Methode zur Identifikation der vorgestellten Aufgaben. Für die weitere Identifikation der Parameter würde allerdings die Partikelschwarmoptimierung herangezogen, da es sich bei dieser um ein globales Lösungsverfahren handelt, wodurch der Identifikationsprozess erleichtert wurde. Diese Methode erfordert, die Definition der Grenzen des Parametersuchraumes und nicht möglichst geeignete Startwerte, die nahe den gesuchten globalen Minimum liegen müssen. Die im Rahmen dieser Arbeit herangezogenen Datensätze konnten hierbei aus [4] und [5] entnommen werden. Diese würden unter der Verwendung von primären und sekundärem Zink-Dialkyl-Dithiophosphat (ZDDP) bei unterschiedlichen Temperaturen und Phosphor-konzentrationen aufgezeichnet.

3. MODELLBILDUNG

Die Modellbildung stützt sich auf geometrische Beobachtungen aus Reibversuchen. In einem System mit leichtem Verschleiß wurde durch ein ZDDP-Additiv ein Tribofilm aufgebaut, der danach durch Reibung unter einem Basisöl ohne ZDDP Zusatz wieder abgetragen wurde. Auch wenn die dabei auftretende Abtragsrate gering war, impliziert dieses Verhalten, dass sich in dem System mit ZDDP ein Gleichgewicht zwischen Wachstum und Abtrag eingestellt hat, welches durch das Fehlen von ZDDP gestört wurde [4]. Es sollte dabei aber gewährleistet sein, dass die Wachstumsrate der Triboschicht größer als die Abtragsrate ist. [6]. Diese Beobachtungen decken sich auch mit jenen aus Messungen aus Transmissionselektronenmikroskopie-Studien (TEM-Studien), die an Partikeln durchgeführt wurden, welche aus Systemen mit milden Bedingungen stammen. Diese bestehen hauptsächlich aus amorphen Teilchen, die chemisch aus den im Schmierstoff vorhandenen Additiven gebildet wurden im Gegensatz zu kristallinen, ausgebrochenen Verschleißpartikeln [7]. Im Allgemeinen bilden Tribofilme Schichtdicken im Bereich von $50 - 200\ nm$ aus, wobei das Schichtwachstum in zwei Stufen abläuft. Zuerst bildet sich eine gleichmäßige Schicht bis zu einer Dicke von $\sim 20\ nm$ und dann Inseln mit einer deutlich größeren Schichtdicken ($\sim 100\ nm$), welche linienförmig entlang der Gleitrichtung angeordnet sind (siehe Abb.1). Diese Inseln wachsen dann über die Zeit sowohl in der Breite als auch in der Höhe. Das Modell bildet hingegen nur unidirektionale Ausdehnung in laterale Richtung ab.

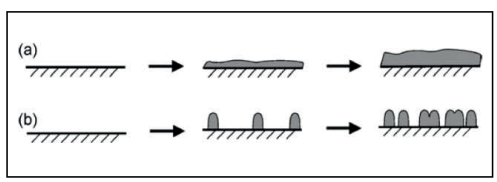


Abb. 1: (a) homogenes und (b) inselartiges Schichtwachstum [37]

3.1 Gleichmäßiges Schichtwachstum

Gleichmäßiges Schichtwachstum tritt nur bis zu einer Schichtdicke von ~ 20nm auf und ist ein erster dünner Film, der vor dem Inselwachstum entsteht. Dieses Verhalten kann man darstellen als die zeitliche Änderung der mittleren Schichtdicke, gleichgesetzt mit einem konstanten Faktor. Die Lösung dieser Differentialgleichung wäre eine lineare Funktion und könnte das Wachstum nur in einem kleinen Ausschnitt ausreichend beschreiben. Hierzu konnte in [8] gezeigt werden, dass dies für einen AFM-Versuch bei einer Last von 20 ± 0.1 nN bei einer Zykluszahl von > 6000 für die ersten ~ 1200 Zyklen zutreffend ist. Dieses Modell würde allerdings bei der Betrachtung dickerer Schichten zu Abweichungen führen.

3.2 Unidirektionales Schichtwachstum

Bei einer Schichtdicke von $> 20\ nm$ kann somit angenommen werden, dass die Schicht homogen anwächst. Hierbei wird davon ausgegangen, dass das Wachstum des Tribofilms beziehungsweise das Wachstum der Inseln proportional zur nicht bedeckten Oberfläche ist. Dieser Zusammenhang lässt sich wie folgt darstellen:

$$\frac{dX}{dt} = k_1(1 - X). \tag{1}$$

Hierbei stellt X den Anteil der mit Tribofilm bedeckten Oberfläche und k_1 die Ratenkonstante für das Schichtwachstum dar. Löst man diese gewöhnliche Differentialgleichung analytisch, so führt dies zu folgendem Ausdruck,

$$X = 1 - e^{-k_1(t-t_0)}. \tag{2}$$

In dieser Gleichung stellt t_0 den Startpunkt der Modellkurve dar. Die mittlere Schichtdicke h_{mean} des Tribofilms berechnet sich dann gemäß

$$h_{mean} = h_{max}X, \tag{3}$$

wobei h_{max} die maximale Filmdicke ist. Zu Beginn des Reibprozesses kann man sich die Inseln mit einer konstanten Höhe von h_{max} regelmäßig über die Oberfläche verteilt vorstellen. Diese wachsen dann gemäß der Wachstumskonstante k_1 in die Breite, bis sie sich berühren (siehe Abb. 2). In diesem Modell wird also nicht zwischen Breiten- und Höhenwachstum unterschieden.

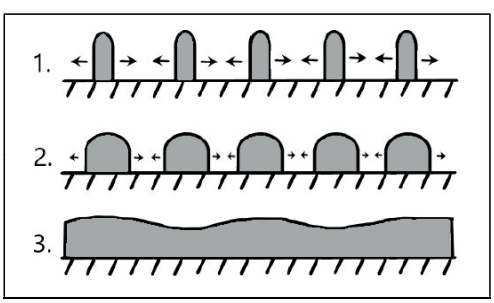


Abb. 2: schematische Darstellung des Modells für unidirektionales Wachstum

3.3 Gleichmäßiges Schichtwachstum

Für die mathematische Beschreibung des Schichtabtrages wird angenommen, dass die Abtragsrate proportional zu einer Funktion der Filmdicke oder -bedeckung ist, wobei eine Potenzbeziehung angenommen wurde,

$$\frac{dX}{dt} = -k_5(X - X_0)^n. \tag{4}$$

Löst man diese Differentialgleichung analytisch, so führt dies zu folgenden Ausdrücken

$$\begin{aligned} n = 1:&\quad X = e^{-k5(t-t_\tau)}(1 - X_0) + X_0 \\ n > 1:&\quad X = (k_5(t - t_\tau)(n-1) + (n-1) + (1 - X_0)^{1-n})^{-\frac{1}{n-1}} + X_0 \end{aligned} \tag{5}$$

wobei X_0 einen nicht entfernbaren Restanteil der Schicht und k_5 die Ratenkonstante für den Abtrag darstellt. Die mittlere Schichtdicke findet man wieder durch (3).

3.4 Kombinierte Ratengleichung

An dieser Stelle werden die Gleichungen (1) und (4) zu einem kombinierten Modell zusammengesetzt. Damit ergibt sich die folgende Differentialgleichung,

$$\frac{dX}{dt} = k_1(1 - X) - k_5(X - X_0)^n. \tag{6}$$

In [37] wurde festgestellt, dass k5 von der Reibzeit abhängt, da der Abtragsprozess erst ab einem gewissen Zeitpunkt einsetzt. Dies lässt sich durch die Beziehung

$$\begin{aligned} t < t_\tau:&\quad k_5 = 0 \\ t > t_\tau:&\quad k_5 = k_{5mx}(1 - e^{-k_6(t-\tau)}) \end{aligned} \tag{7}$$

ausdrücken, wobei k_6 und k_{5mx} Gewichtungsfaktoren darstellen. Setzt man nun (7) in (6) ein so ergibt sich

$$\frac{dX}{dt} = k_1(1 - X) - k_{5mx}(1 - e^{-k6(t-\tau)})(X - X_0)^n. \tag{8}$$

Um diese Differentialgleichung lösen zu können, muss auf ein numerisches Lösungsverfahren zurückgegriffen werden, da die Gleichung (8) nicht analytisch lösbar ist. Die Differentialgleichung zur Berechnung der mittleren Schichtdicke h_{mean} lautet nun

$$h_{mean} = h_{max}(k_1(1 - X) - k_{5mx}(1 - e^{-k_6(t-\tau)})(X - X_0)^n). \tag{9}$$

3.5 Die spannungsunterstützte thermische Aktivierung - das Arrheniusmodell

Der Ansatz basiert auf der Boltzmann-Verteilung, die die Aufteilung von Teilchen oder Molekülen auf quantisierten Energieniveaus darstellt. Diese beschreibt die Besetzung der Energiezustände im thermodynamischen Gleichgewicht E_i, wobei die Besetzungszahl n einer Exponentialfunktion gehorcht, $T[K]$ die absolute Temperatur und $R = 8,3145\ \frac{J}{molK}$ die Gaskonstante bezeichnet [5],

$$n(E_i) \propto e^{-\frac{E_i}{RT}}. \tag{10}$$

Hierbei lässt sich die Reaktionszeit k_1, mit der Teilchen chemische Zustände durchlaufen, durch den von Svante Arrhenius beschriebenen Zusammenhang zwischen Bolzmannverteilung und der Aktivierungsenergie E darstellen, wobei A ein Vorfaktor ist,

$$k_1 = A\, e^{-\frac{E}{RT}}. \tag{11}$$

Geht man nun von der Überlegung des spannungsinduzierten thermischen Aktivierungsmodells aus, so lässt sich festhalten, dass eine aufgebrachte Last auf atomarer Ebene zu einer Reduktion der Aktivierungsenergie in Vorwärtsrichtung führt

(siehe Abb.3). Bindet man diese Erkenntnis formal in (11) ein, so erhält man folgende Gleichung:

$$k_1 = A\, e^{-\frac{E - f\Delta x}{RT}}. \quad (12)$$

Der Ausdruck Δx ist die Aktivierungsweglänge, die ein Teilchen zurücklegen muss, um die erforderliche Arbeit $f\Delta x$ zu verrichten, sodass dieses seinen energetischen Zustand ändern kann. Dieser Ausdruck kann auch auf molare Größen hoch skaliert werden, wodurch folgender Ausdruck zustande kommt:

$$k_1 = A\, e^{-\frac{E - F\Delta x}{RT}}. \quad (13)$$

k_1 stellt in dieser Darstellung nur die Vorwärtsgeschwindigkeitskonstante des Prozesses dar und E ist die molare thermische Aktivierungsenergie.

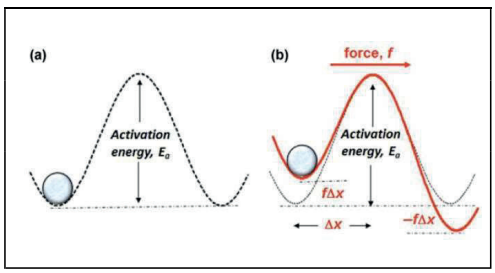


Abb. 3: (a) schematische Darstellung der Aktivierungsenergiebarriere ohne die Einwirkung einer Kraft
(b) Einfluss einer angelegten Kraft auf die Energiebarriere

3.6 Die spannungsunterstützte thermische Aktivierung - Mechanoaktivierung

Chemische Reaktionen sind im allgemeinen aktivierte Prozesse, die einen gewissen Energieeintrag zur Überwindung der Potenzialbarriere benötigen. Bei diesem Energieeintrag handelt es sich um die Aktivierungsenergie des Prozesses, wobei es sich bei thermoaktivierten Reaktionen um einen Wärmeeintrag und bei mechanoaktivierten Reaktionen um eine mechanisch eingebrachte Energie, handelt.

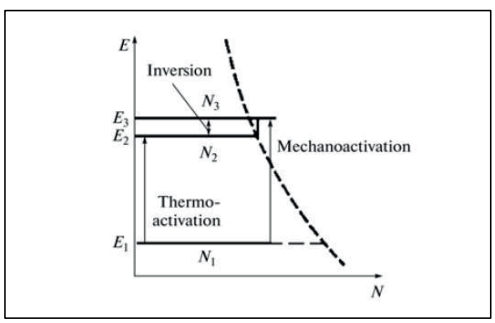


Abb. 4: Inverse Population im Dreistufensystem mit dem Grundniveau E_1, dem thermisch aktivierten Niveau E_2, dem mechanoaktivierten Niveau E_3 und der zugehörigen Anzahlen an Teilchen $N_1 - N_3$ [41].

Die Entstehung von Triboschichten folgt, einer tribochemischen Reaktion, die durch den Reibprozess aktiviert wird. Somit ist die Triboaktivierung eine Kombination aus Mechano- und Thermoaktivierung. Hierzu konnte in [9] der Hauptunterschied zwischen tribologisch aktivierten und thermisch aktivierten chemischen Reaktionen hinsichtlich deren Population der Übergangs-zustände herausgearbeitet werden, indem diese bei ähnlichen Energieniveaus gegenübergestellt wurden. Dieser besteht darin, dass triboaktivierte Prozesse eine umgekehrte Population an Übergangszuständen aufweisen, als in der Physik üblich. Denn die Besetzungsinversion beschreibt den Zustand eines Systems, in dem sich mehr Teilchen auf einem energetisch höheren Niveau befinden als dem energetisch niedrigeren. Dies ist bei thermisch aktivierten chemischen Prozessen im Wärmegleichgewicht der Fall, bei triboaktivierten Prozessen jedoch genau invers (siehe Abb.4). Die Teilchenverteilung im thermodynamischen Gleichgewicht wird durch die gestrichelte Linie dargestellt. Die Population der Übergangszustände bezeichnet die Gesamtheit aller Teilchen, die sich im jeweiligen Zustand befinden. In [10] wird hierzu folgende Geschwindigkeitskonstante eingeführt, die diese Umwandlung beschreibt

$$k_{tribo-thermo} = \frac{x_{tribo}}{x_{thermo}} k_{thermo}. \quad (14)$$

Hierbei repräsentiert x_{tribo} und x_{thermo} die Anteile der Population des Übergangszustandes aufgrund von thermischer Aktivierung bzw. Mechanoaktivierung im Vergleich zum Grundzustand des Systems und k_{thermo} die thermische Geschwindigkeitskonstante, dir für einen aktivierten Komplex wie folgt modelliert werden kann,

$$k_{thermo} = \frac{kT}{h} e^{-\frac{\Delta H}{RT}}. \quad (15)$$

Des Weiteren stellt k die Bolzmannkonstante, h die Plank's Konstante und ΔH ist die Aktivierungsenthalpie dar und x_{thermo} kann wie folgt dargestellt werden

$$x_{thermo} = e^{-\frac{\Delta H}{RT}}. \quad (16)$$

Setzt man an dieser Stelle die Gleichungen (15) und (16) in die Gleichung (14) ein, so kann folgender Ausdruck für die Ratenkonstante berechnet werden

$$k_1 = k_{tribo-thermo} = \frac{kT}{h} x_{tribo}. \quad (14)$$

4. PARAMETERIDENTIFIKATION

Im Rahmen der Parameteridentifikation wurden die unbekannten Größen der vorgestellten Modelle (Gleichung (1), (4) und (8)) für unterschiedliche Datensätze identifiziert. Die unter 3.5 und 3.6 vorgestellten Modelle für die Ratenkonstante des Wachstumsterms wurden gleichermaßen zusammen mit dem Wachstumsmodell bestimmt.

4.1 Das Wachstumsmodell

Für die Betrachtung des Schichtaufbaues wurden Datensätze herangezogen, die aus Reibversuchen stammen, die mit primären ZDDP durchgeführt wurden. Die Idee hinter dieser Herangehensweise ist, dass bei kontinuierlich anwachsender Schicht, bis zum Erreichen der stationären Schichtdicke, der Schichtdickenwert nicht abnimmt. Dies ist bei anderen Additivzusammensetzungen nicht zwangsläufig der Fall. Abbildung (5) zeigt exemplarisch den Ergebnisplot zu einem Versuch mit primärem ZDDP (0.08%P) bei 50°C.

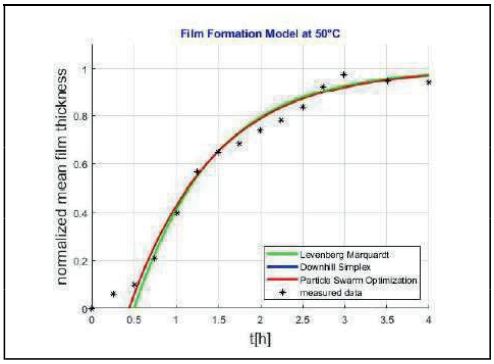


Abb. 5: Gegenüberstellung der berechneten Ergebnisse auf Basis des Levenberg-Marquardt Verfahrens, des Downhill-Simplex-Algorithmus und der Partikelschwarmoptimierung bei 50°C

Abbildung (6) zeigt, dass die Ratenkonstante die Steigung der Tangente zu Prozessbeginn repräsentiert, wodurch diese als Startgeschwindigkeit des Teilprozesses verstanden werden kann.

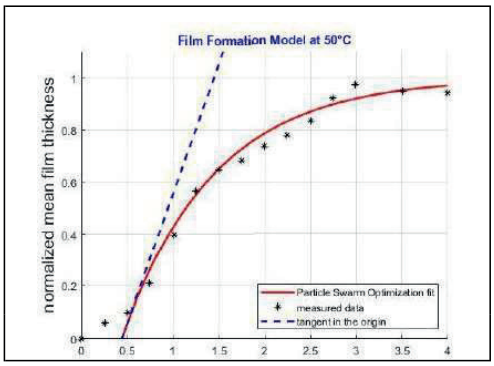


Abb. 6: Ergebnisse der Partikelschwarmoptimierung des Wachstumsmodells und die Tangente im Ursprung bei 50°C

4.2 Das Abtragsmodell

Eine separierte Betrachtung des Schichtabtrages anhand der dafür vorgesehenen Testdaten kann ebenso wie die isolierte Betrachtung des Schichtaufbaues nur theoretisch vorgenommen werden. Hierzu wurde zur Generierung der Testdaten ein Szenario geschaffen, wo der Abtrag durch Veränderung der Basisölmischung herbeigeführt wurde. Es wurde nach Erreichen der maximalen Schichtdicke der Abtrag durch Ersetzen der additivhaltigen Ölmischung mit Basisöl+Dispergiermittel eingeleitet. Auch für die hierzu gehörigen Datensätze konnten gut Ergebnisse erzielt werden. Hierzu zeigt Abbildung (7) exemplarisch den Ergebnisplot zu einem Versuch zum Abtrag der Schicht mit primärem ZDDP (0.08%P) bei 50°C. Auch für den Teilprozess des Schichtabtrages entspricht die Ratenkonstante der negativen Prozessgeschwindigkeit zu Beginn des Teilprozesses.

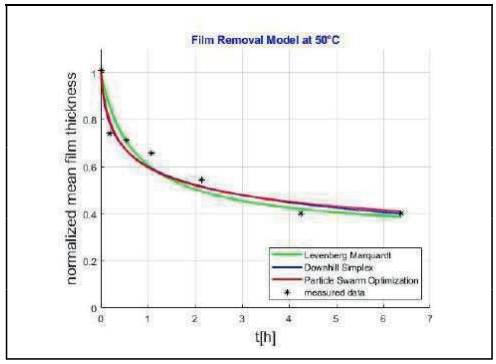


Abb. 7: Gegenüberstellung der berechneten Ergebnisse auf Basis des Levenberg-Marquardt Verfahrens, des Downhill-Simplex-Algorithmus und der Partikelschwarmoptimierung für den Abtrag.

4.3 Das Kombinierte Modell

An dieser Stelle werden die Ergebnisse für die Datensätze die durch Reiben mit primären und sekundären ZDDP bei unterschiedlichen Temperaturn und Phosphorkonzentrationen aufgezeigt. Auch hier zeigt sich gute Übereinstimmung.

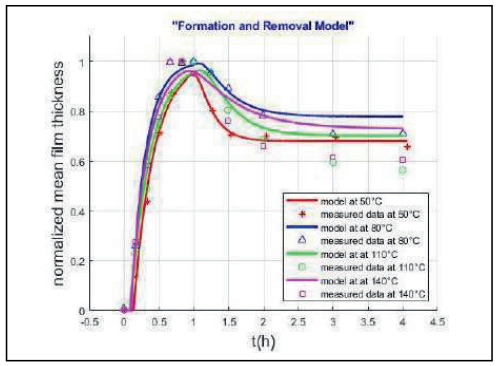


Abb. 10: Darstellung der Ergebnisse für unterschiedliche Temperaturen (sekundäres ZDDP)

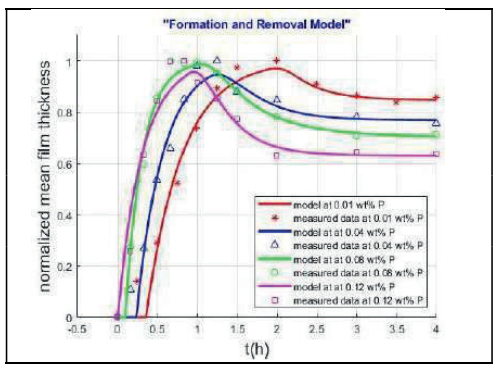


Abb. 11: Darstellung der Ergebnisse für unterschiedliche Phosphorkonzentrationen (sekundäres ZDDP)

4.4 Die spannungsunterstützte thermische Aktivierung - das Arrheniusmodell

Aufgrund von unzureichenden Daten ist die Identifikation des Arrheniusmodells leider nicht gelungen. Alternativ kann die Aktivierungsenergie B gemäß spannungsunterstützter thermischer Aktivierung mit Hilfe der identifizierten Werte der Wachstumsratenkonstante k_1 und der Berechnungs-vorschrift

$$B = -\ln\left(\frac{k_1}{A}\right) RT, \tag{15}$$

durchgeführt werden.

4.5 Die spannungsunterstützte thermische Aktivierung - Mechanoaktivierung

Im folgendem wird der Ansatz der Mechanoaktivierung (Glg.(14)) in das Wachstumsmodell (Glg.(1)) eingesetzt. Daraus ergibt sich folgende Differentialgleichung

$$\frac{dX}{dt} = \frac{k_1}{h} x_{tribo}(1 - X). \tag{1}$$

Die gesuchten Größen sind hierbei x_{tribo}, der Anteile der Population des Übergangszustandes aufgrund von Mechano-aktivierung und t_0 der Startzeitpunkt. Hierzu könnten folgende Werte gefunden werden.

primäres ZDDP						
Nr	condition	xtribo	t0	R^2	RMSE	elapsed time (s)
1	50°C	1.6278e-13	0.438	0.9883	0.0374	61.50
2	80°C	1.8422e-13	0.470	0.9979	0.0157	46.50
3	110°C	1.5509e-13	0.446	0.9926	0.0330	58.30
4	140°C	1.4324e-13	0.295	0.9805	0.0528	92.76

Abb. 12: Die identifizierten Parameter zu 4.5

Anhand dieser Tabelle ist erkennbar, dass x_{tribo} annähernd konstant erscheint. Diese Annahme lässt sich anhand der folgenden Abbildungen (13) und (14) bestätigen. Abbildung (13) zeigt die Ergebnisplots mit identifizierten Werten und Abbildung (14) zeigt die Ergebniskurven für x_{tribo}= 1.6133e-13.

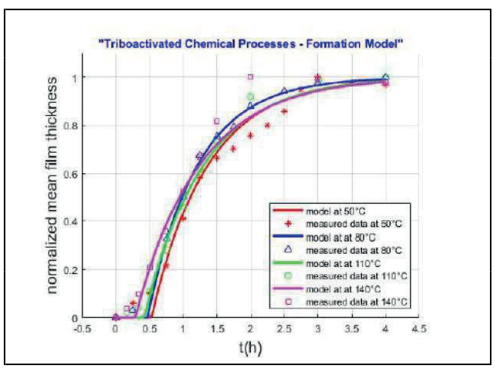


Abb. 13: Darstellung der Ergebnisse für unterschiedliche Phosphorkonzentrationen (sekundäres ZDDP)

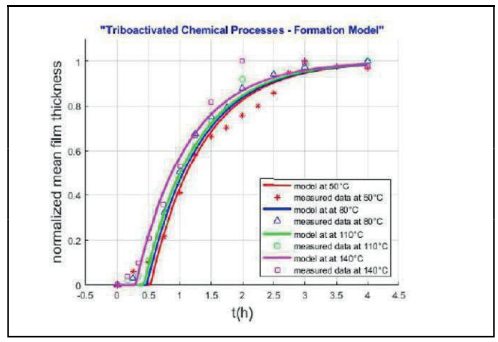


Abb. 14: Darstellung der Ergebnisse für unterschiedliche Phosphorkonzentrationen (sekundäres ZDDP)

4.5 Darstellung der Abhängigkeit der Wachstumsratenkonstante

Die folgenden Abbildungen (34) und (35) zeigen die Wachstumsratenkonstanten in Abhängigkeit der Kontakt-temperatur und der Phosphorkonzentration, jeweils für primäres ZDDP und sekundäres ZDDP. Da je nur acht Datensätze für Kontakte bei $50, 80, 110{,}140°C$ bei $0,08wt$, und $0.01, 0.04, 0.08, 0.12wtP$ bei $80°C$ vorhanden sind, wurden die restlichen Werte, rechnerisch ermittelt. Daher ist es wichtig zu erwähnen, dass diese Grafiken lediglich eine Übersicht darstellen und hinsichtlich Genauigkeit und Richtigkeit durch weitere Messungen verbessert bzw. überprüft werden müssen.

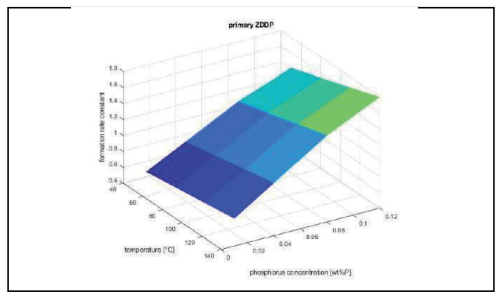


Abb. 15: Zusammenhang zwischen der Phosporkonzentration, der Temperatur und der Wachstumsratenkonstante (primären ZDDP)

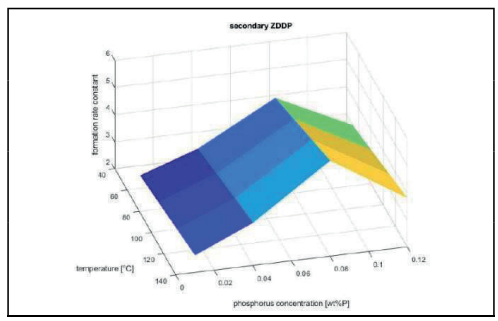


Abb. 13: Zusammenhang zwischen der Phosporkonzentration, der Temperatur und der Wachstumsratenkonstante (sekundäres ZDDP)

5. DISKUSSION DER ERGEBNISSE

5.1 Betrachtung der Teilmodelle

Das Wachstumsmodell als solches ist nicht in der Lage, die Schichtbildung über den gesamten Zeitbereich abzubilden, da die ersten 30 min des Reibprozesses nicht dargestellt werden können. In diesem kurzen Abschnitt wächst die Schicht näherungsweise linear an. In [8] wurde vorgeschlagen, dass es sich hierbei um eine Reaktion nullter Ordnung handelt. Danach lässt sich die Änderung der Schicht durch eine Exponentialfunktion beschreiben, bis die maximale Schichtdicke erreicht wird und das Wachstum stagniert. Dieser Bereich kann vollständig durch das Teilmodell des Wachstums dargestellt werden. Auch der Abtragsterm ist dazu geeignet die gemessenen Zeitreihen für den Abtrag darzustellen. Die Ratenkonstante des Wachstumsmodells und des Abtragsmodells entsprechen der Steigung der Tangente zu Teilprozessbeginn. Hieraus lässt sich ableiten, dass es sich bei der Ratenkonstante um die Prozessgeschwindigkeit handelt und es gilt, $r_{Wachstum_0} = k_1$ und $r_{Abtrag_0} = -k_5$.

5.2 Betrachtung des zusammengesetzten Modells

Das Modell wurde hierfür auf alle vorhandenen Datensätze (primäres und sekundäres ZDDP) angewendet. Für primäres ZDDP hat sich herausgestellt, dass das kombinierte Modell für die Wachstumsparameter, dieselben Ergebnisse wie das Teilmodell liefert. Da im betrachteten Zeitraum kein Abfall der Schichtdickenwerte gemessen wurde, ist der Abtragsterm unwirksam, wodurch die gefundenen Parameter für den Abtrag keinerlei Aussagekraft haben. Die Ergebnisse des Modells für die Datensätze mit sekundären ZDDP führen zu einem plausiblen Ergebnis, da die Schichtdicke ab einem Zeitpunkt τ abfällt.

5.3 Modellierung der Wachstumsratenkonstante

Tribofilme sind tribochemisch aktivierte Prozesse, die eine mechanische Aktivierung erfordern, damit sich ein robuster Verschleißschutz ausbilden kann. Im Rahmen dieser Arbeit konnte die Ratenkonstante des Wachstumsterms so modelliert werden, dass die Kontakttemperatur abgebildet werden kann. Es wurde zunächst versucht, diese mit Hilfe der Arrheniusgleichung zu bestimmen, was allerdings nicht zum gewünschten Erfolg geführt hat, da zu wenig Messwerte zur Verfügung standen. Des Weiteren wurde in der Literatur beschrieben, dass die Anwendung der Arrheniusgleichung auf triboaktivierte Prozesse zu ungewöhnlich niedrigen Aktivierungsenergien führt. Es wurde ein weiterer Ansatz zur Modellierung der Wachstumsrate eingeführt, welcher auf dem Verhältnis der anteiligen Populationen der Übergangszustände auf Basis der thermischen bzw. mechanochemischen Aktivierung und der Geschwindigkeits-konstante für die chemische Reaktion beruht. Dieser liefert zufriedenstellende Ergebnisse. Des Weiteren konnte festgestellt werden, dass die Abhängigkeit der Phosphorkonzentration jedenfalls einen größeren Einfluss als die Temperatur auf die Filmbildung hat.

6. ZUSAMMENFASSUNG UND AUSBLICK

Im Rahmen dieser Arbeit wurde ein Modell, für den Aufbau und Abtrag von Triboschichten erarbeitet, die auf den Erkenntnissen aus tribologischen Experimenten basieren. Die Basismodelle und die verwendeten Datensätze wurden aus einer bereits in der Literatur vorhandenen Arbeit entlehnt, analysiert und anschließend weiterentwickelt. Die Identifikation der Parameter wurde überwiegend mit der Partikelschwarmoptimierung durchgeführt. Des Weiteren konnten in allen Bereichen der Arbeit gute Ergebnisse erzielt werden. Da mit diesem Modell bisher nur die Temperaturabhängigkeit abgebildet werden kann, ist es im Rahmen von weiterer Projekttätigkeit von Interesse auch andere Einflussgrößen zu analysieren und zu modellieren.

7. LITERATURVERZEICHNIS (10 PT, CAPITALS, BOLD)

[1] J. Lenard. An examination of the coeffcient of friction, pages 85-114. 12 2002. doi:10.1016/B978-008044024-8/50006-0.

[2] A. Morina und A. Neville. Tribofilms: aspects of formation, stability and removal. Journal of Physics, 2007. doi:10.1088/0022-3727/40/18/s08.

[3] J. Zhang und H. A. Spikes. On the mechanism of zddp antiwear film formation. Tribology Letters, 2016.

[4] H. Fujita, R. P. Glovnea, und H. A. Spikes. Study of zinc dialkydithiophosphate antiwear film formation and removal processes, part i: experimental. Tribology Transactions, 48:558 _ 566, 2005.

[5] H. Fujita und H. A. Spikes. Study of zinc dialkyldithiophosphate antiwear film formation and removal processes, part ii: kinetic model. Tribology Transactions, 2005, doi:10.1080/05698190500385187.

[6] R. Bosman und D. J. Schipper. Mild wear prediction of boundary-lubricated contacts. Tribology Letters, 2011. doi:10.1007/s11249-011-9760-3.

[7] Y. R. Li, G. Pereira, M. Kasrai, und P. R. Norton. Studies on zddp anti-wear films formed under different conditions by xanes spectroscopy, atomic force microscopy and 31p nmr. Tribology Letters, 2007, doi:10.1007/s11249-007-9275-0.

[8] N. N. Gosvami, J. A. Bares, F. Mangolini, A. R. Konicek, D. G. Yablon, und R. W. Carpick. Mechanisms of antiwear tribofilm growth revealed in situ by single-asperity sliding contacts. Science Express, 2015. doi:10.1126/science.1258788.

[9] S. B. Bulgarevich, M. V. Boiko, V. I. Kolesnikov, und K. E. Korets. Population of transition states of triboactivated chemical processes. Journal of Friction and Wear, 2010.

[10] S. Bulgarevich, M. Boiko, V. Kolesnikov, und V. Feizova. Thermodynamic and kinetic analyses of probable chemical reactions in the tribocontact zone and the effect of heavy pressure on evolution of adsorption processes. Journal of Friction and Wear 2011, doi:10.3103/S1068366611040027.



Viktoria Seidl, B.Sc.

AC2T research GmbH, Viktor Kaplan-Straße 2/c, +43 262281600316, viktoria.seidl@ac2t.at

2021- Wissenschaftlicher Mitarbeiter bei AC2T research GmbH
2019-2023 Masterstudiengang Mechatronik Fachhochschule Wiener Neustadt
2016-2019 Bachelorstudiengang Mechatronik | Mikrosystemtechnik Fachhochschule Wiener Neustadt

A PCB-BASED APPROACH TOWARDS HIGH-SPEED DATA TRANSMISSION AND POWER MANAGEMENT FOR THE BCM' SYSTEM INSIDE THE ATLAS EXPERIMENT OF THE LHC

Swoboda A.

Abstract: *The CERN LHC consists out of four major experiments, one of them is the ATLAS experiment, which first discovered the existence of the Higgs Boson. Inside the LHC two particle beams are accelerated to almost speed of light, before they collide at the locations of the different experiments.*
Since the maximum energy of the particle bunches is 13 TeV, with the upgrade to HL-LHC the collision energy rises to up to 14 TeV, abnormal beam conditions are an immense threat to the ATLAS experiment and its different detectors and trackers.
The BCM' system is the follow-up of the previous BCM system, which will be implemented in the HL-LHC upgrade. Two detector stations constantly monitor the particle beam and can cause a beam dump, if severe damage to the ATLAS experiment, respectively the LHC, is given.

This thesis is covering the development of a proof-of-concept for the readout and powering of the BCM' module rings. A PCB is developed, which powers the BCM' modules, routes the high-speed data of 1.28 Gb/s towards the backend-readout and monitors the voltages and temperatures of the modules.

Simulations and calculations for proper LVDS data transmission are done, and a test-setup close to the readout chain inside the ATLAS experiment is developed.

Keywords: *LVDS, differential data transmission, ATLAS experiment, beam conditions monitoring, transmission lines*

1. INTRODUCTION

At the core of the research done at CERN is the Large Hadron Collider (LHC), which is a particle accelerator with a total circumference of 27 km, used for colliding protons and heavy ions. The research done at CERN helps discovering the secrets of our universe, with the goal of getting a broader understanding of the matter our world is made of as well as the dark matter, for which scientists search meticulously.
Four big experiments are used to conduct physics at CERN, one is the ATLAS experiment, a particle detector of 44 m in length and 25 m in height. This detector is used as a general purpose detector, contributing with its measurements to a broad range of physics experiments.
The ATLAS experiment hosts different detector systems, which are arranged in different layers around the beam pipe [ATLAS collaboration 2008].
Directly around the beam pipe a safety system is placed, the ATLAS beam conditions monitoring (BCM) system, which is capable to detect abnormalities in the beam circulating in the LHC beam tube and has the ability to call for an abort, which leads to a beam dump, if severe damage to the LHC infrastructure is apprehended [Pernegger 2005].
The upgrade of the LHC to high-luminosity, which increases the current design luminosity of the LHC, $10^{!"}cm^{\#\$}s^{\#\%}$, by a factor of ten. More luminosity means more bunch crossings and therefore more collisions, which gives more data to study.
Higher luminosity also exposes the detector to more radiation, which calls for a new BCM system.
This new beam monitoring system called BCM', uses a front-end readout ASIC, which provides digitation of the signals at the front end. The heart of the BCM' system consists out of the BCM' modules, which fulfill two different purposes:

- -monitoring of the luminosity
- -monitoring of potential beam abnormalities which call for an abort

The BCM' modules therefore are split into two parts, one for luminosity measurement and one for triggering and abort, if necessary. The modules themselves consist out of polycrystalline chemical vapor deposition (pCVD) diamond sensors, a calypso ASIC to digitize the signals and a driver ASIC, the so-called LAPA, which boosts the signals from Calypso.
The BCM' modules are placed on two rings, each holding 4 modules in a 90-degree spacing and a distance between the two rings of z = 1.9 m [Dolenc et al. 2008].

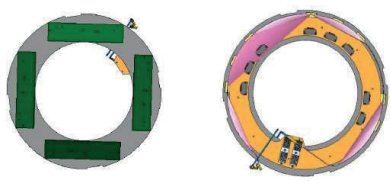


Figure 1: Rendering of the BCM' stations. **Left:** The BCM' ring with the modules attached. **Right:** The aPCB, which will be connected via flexes to the module ring.

In order to read the signals from the BCM' modules and to power and program them, a second ring-shaped PCB comes into play. This PCB is the so-called auxiliary PCB (aPCB), which is connected to the ATLAS services for power and communication. aPCB will power and read out four modules, which means that two aPCBs will be needed in the whole BCM' system. The signal speed on aPCB is 1.28 Gb/s. The connection between aPCB and the module ring is established via flex-PCBs.
This work is about developing a proof-of-concept PCB, which should mimic the functionality of aPCB.

2. PROBLEM AND TASK

The main problem with the development of aPCB is the availability of the so-called PP0 to Type0 flex, which is a flex PCB that is connected to the central powering and readout services of the ATLAS experiment. Therefore a proof-of-concept PCB has to be developed, which does not need this PP0 to Type0 flex.
This PCB is the so-called xPCB, which will fulfill the same functionality as aPCB, but only serves half the modules. The connectors on xPCB are also slightly different than those used

on aPCB, in order to only use components able to be purchased and used.
The biggest block of xPCB is to route the high-speed signals coming from the BCM' modules to the output connectors. The signals coming from the front-end ASIC, Calypso, correspond to the LVDS standard, which calls for precise impedance matching throughout the whole transmission chain.
The BCM' modules host polycrystalline chemical vapor deposition (pCVD) diamond sensors, which require high voltage of 1 kV in order to be operated. This high power needs to be routed over xPCB as well.
The Calypso and the LAPA require low voltage supplies of 1.2 V, respectively 1.8 V. The ATLAS power services only offer 2.5 V, therefore step-down buck converters are placed on the module ring. Those buck converters are the so-called bPOLs, which are low-noise DC-DC converters developed at CERN for converting voltages in high radiated environments [CERN DCDC project 2023].
Calypso and LAPA are programmable via I2C, those lines also need to be placed on xPCB.
The main task is to develop this xPCB and test it thoroughly, with a readout chain as close as possible to the final system.

3. MATERIALS AND METHODS

3.1 Simulations of LVDS lines

To understand the importance of correct impedance matching, simulations are performed. The software used for this task is LT-spice.
Transmission lines are viewed as distributed parameter networks, which means the voltages and currents can vary in magnitude and phase throughout the length of the transmission line. The main parameters of transmission lines are their characteristic impedance, Z_0, and the propagation speed of the signal, v.
The telegrapher's equations help us finding the current and the voltage on a transmission line any time *t* and location *x*.

$$\frac{di(x,t)}{dx} = -G \cdot v(x,t) - C \cdot \frac{dv(x,t)}{dt} \quad (1)$$

$$\frac{dv(x,t)}{dx} = -R \cdot i(x,t) - L \cdot \frac{di(x,t)}{dt} \quad (2)$$

With R being the series resistance per unit length of line, being the series inductance per unit length of line, G the shunt conductance per unit length of line and C being the shunt capacitance per unit length of line [Pozar D.M. 2011].
For the derivation of the characteristic impedance equations (1) and (2) are excitated with harmonic, sinusoidal signals:

$$Z_0 = \sqrt{\frac{Z}{Y}} = \sqrt{\frac{R+j\omega L}{G+j\omega C}} = \sqrt{\frac{L}{C}} \cdot \frac{1 - j\frac{R}{\omega L}}{1 - j\frac{G}{\omega C}} \quad (3)$$

For frequencies above 10 kHz $j\frac{R}{\omega L} \ll 1$ and $j\frac{G}{\omega C} \ll 1$.
Therefore Z_0 reduces to:

$$Z_0 = \sqrt{\frac{L}{C}} \quad (4)$$

This simplification does not mean that we do not have to deal with losses in case of high frequencies. On the contrary, we must deal with two different types of losses, friction losses in the dielectric and ohmic losses due to the skin effect.
The voltage on a transmission line consists out of two waves, an incident wave and a reflected wave. The amplitude of the reflected wave is highly dependent on the termination of the transmission line, and the reflection coefficient ρ is given by:

$$\Gamma = \frac{V_0^-}{V_0^+} = \frac{Z_L - Z_0}{Z_L + Z_0} \quad (5)$$

With V_0^- being the reflected voltage wave and V_0^+ the incident voltage wave.
For the study of the behavior of the transmission line an LT-Spice schematic was built, consisting out of a transmission line with a delay time of 1ns, which corresponds to the length of the transmission line on aPCB and xPCB.
The input signal was a random digital bitstream, with an amplitude ranging from 0 V to 2.4 V. The clock frequency is 1.28 Gb/s.
Three different scenarios were simulated:

- The input bitstream is at a constant rate, but the characteristic impedance varies throughout the length of the trace:

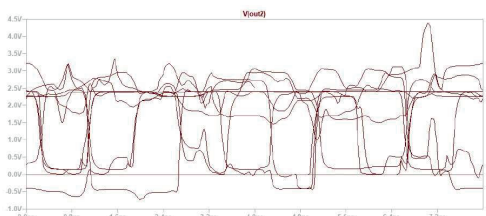


Figure 2: Eye diagram of a transmission line with varying characteristic impedance

- The input bitstream and the characteristic impedance are constant, but the termination resistance varies:

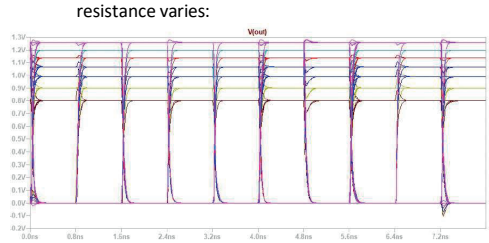


Figure 3: Simulation of the transmission line with a varying termination resistance

The waveform occurs with a different offset for each value of the termination resistance. This phenomenon is called

mismatch loss and described via the Standing wave ratio (SWR):

$$SWR = \frac{V_{max}}{V_{min}} = \frac{1+|\Gamma|}{1-|\Gamma|} \quad (6)$$

- The input bitstream is increasing steadily, which means the frequency of a bit to be a logic 1 is increasing, and the characteristic impedance shows a small change:

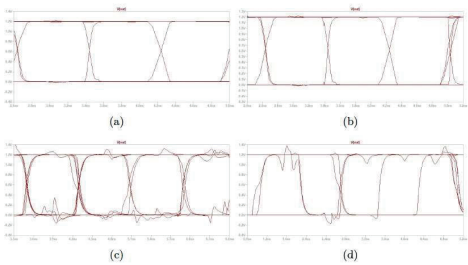


Figure 4: Eye diagrams of the transmission line with different input bit streams. (a): 10% of the transmitted bits are logic 1, (b): 30%, (c):50%, (d):90%.

4. PRACTICAL IMPLEMENTATION

The calculation of the trace width and spacing of xPCB is a highly non-trivial task, which can only done by field solvers. The chosen layer assignment for xPCB is:

Layer	Assignment
Layer 1	LVDS shield / LV ground
Layer 2	LVDS data lines
Layer 3	LVDS shield / HV ground
Layer 4	High voltage
Layer 5	Low Voltage
Layer 6	Rest of Signals

Table 1: Layer assignment on xPCB

Given those assignments, the following design constraints were calculated:

Signal Layer	Shield Layer 1	Shield Layer 2	Trace width [µm]	Trace spacing [µm]
2	3	1	150	142
3	2	4	145	255
4	3	5	150	255

Table 2: Width and spacing of the differential pair traces in the PCB layout

The six-layer xPCB was designed using Altium, the final schematics of all layers of the PCB can be seen in Figure 5.

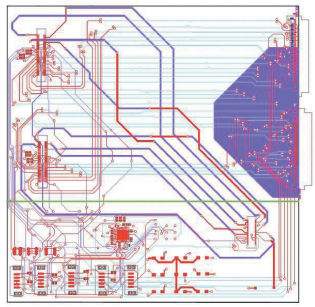


Figure 5: Schematics of all layers of xPCB

5. RESULTS

The testing of xPCB is done with a readout chain, which is as close as possible to the one used in the actual BCM' system.

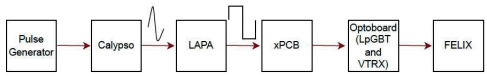


Figure 6: Flow diagram of the readout chain for the xPCB system tests

In a first measurement the accuracy of the transmission was tested directly on xPCB, with the signal propagating a distance of 300 mm. The signal was transmitted properly, which can be seen in the eye-diagram shown in figure 7.

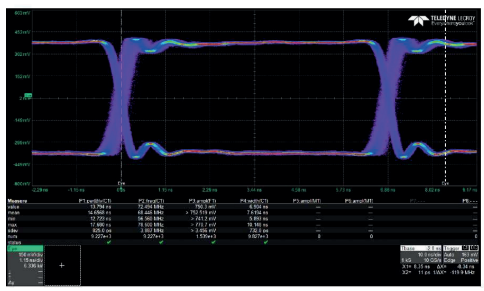


Figure 7: An eye-diagram of the transmitted data over xPCB

In the next measurement the data was read out via the FELIX read-out cards, which are readout-cards delicately developed for the ATLAS experiment. The data coming from the BCM' modules is processed there, the ToA and ToT information is extracted and stored into 32 bit words. FELIX is operated with the ATLAS internal bunch crossing clock, which means that the data processing is synchronized to the ATLAS events [Trovato M. 2019]. One readout cycle in this test consists out of 800 000 incoming bits recorded by FELIX. The ToA and ToT information is then plotted into histograms.

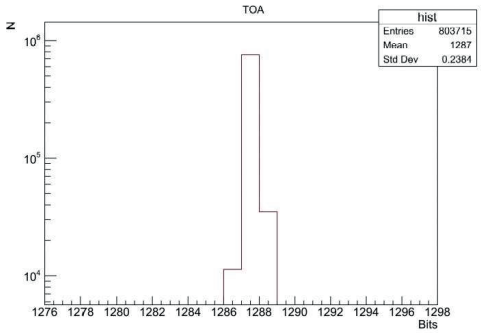


Figure 8: ToA information of the transmitted bits over xPCB, sampled with FELIX

The ToA information in this histogram shows a clear sampling and proper timing over xPCB, every bit sent over xPCB was properly sampled on FELIX side.

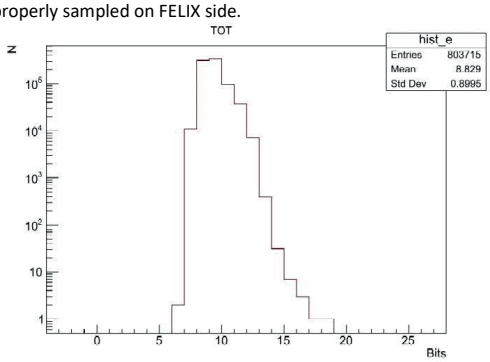


Figure 9: ToT information of transmitted bits over xPCB, sampled with FELIX

The ToT measurements show a similar pattern, with xPCB not decreasing the signal quality throughout the transmission chain.

6. CONCLUSION AND OUTLOOK

The transmission of the high-speed data coming from the BCM' modules can be done via xPCB, properly integrated in the readout chain, which paves the way for the design of aPCB, which will use the same layer stack as xPCB does.
In the next steps, an adapterbaord needs to be developed, in order to be able to fully stress-test xPCB and to use all data channels available.
If the BCM' modules are not available until then, an FPGA needs to be programmed in order to simulate the full module behaviour.
Regarding the functional tests, the next steps are the testing of the communication lines on xPCB, and different grounding and shielding options, which can be done via jumpers placed on xPCB.
The last step of testing xPCB is testing the HV-lines, which will power the pCVD diamond sensors on the modules.
All the future findings will be included in the design of the aPCB, which should be placed inside the detector by December 2024.

7. LITERATURE

ATLAS collaboration, 2008. The ATLAS experiment at the CERN large hadron collider. *Journal of Instrumentation.*

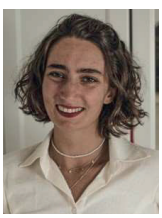
Pernegger, H., 2005. First test results of a high-speed beam conditions monitor for the ATLAS experiment. *IEEE transactions on nuclear science, 52 (5).*

Dolenc et al., 2008. ATLAS diamond beam conditions monitor. *17th international workshop on Vertex detectors.*

CERN, 2023. DCDC project. *https://espace.cern.ch/project-DCDC-new/_layouts/15/start.aspx#/.*

Pozar, D.M., 2011. Microwave engineering. *Wiley-IEEE press, 4th edition.*

Trovato, M., 2019. FELIX: The New Readout System for the ATLAS detector.



Anna Swoboda

CERN, EP-ADE-TK, anna.swoboda@cern.ch

AHS-Matura 2017, Bachelor-Studium „Mechatronik und Mikrosystemtechnik" an der FH Wiener Neustadt 2018 bis 2021, Master-Studium „Mechatronik" an der FH Wiener Neustadt seit 2021

UNTERSUCHUNG EINES KÜHLKÖRPERS ZUR VERBESSERUNG DER KÜHLLEISTUNG

Thalhammer T.

Kurzfassung: Die vorliegende Arbeit befasst sich mit der Untersuchung eines Kühlkörpers, mit dem Ziel dessen Kühlleistung zu verbessern und gleichzeitig den Platzbedarf des Kühlkörpers zu reduzieren. Es sollen auch Möglichkeiten zur Verbesserung der Wartbarkeit und der strömungsmechanischen Energieverluste bei besserer Kühlleistung gefunden werden.
Diese Arbeit beinhaltet theoretisches Wissen der zur Durchführung relevanten Themengebiete. Es folgt eine Erläuterung und Visualisierung der Problemstellung bezüglich der Geometrie des Kühlkörpers und eine detaillierte Beschreibung der Zielsetzung dieser Arbeit. Danach wird auf Basis der numerischen Strömungsmechanik und der Thermodynamik eine Kühlkörpergeometrie untersucht, welche häufig zur Kühlung für die Elektronik diverser Stromrichter zum Einsatz kommt. Um einen realen Belastungsfall darzustellen, werden auf dem Kühlkörper Stromrichtermodule angebracht, welche sich durch eine bestimmte Verlustleistung erhitzen. Anschließend wird eine Parameterstudie mit dem Ziel durchgeführt, Möglichkeiten zu ermitteln, um eine ähnliche oder bessere Kühlleistung mit einer anderer Finnengeometrie, sowie einer besseren Wartbarkeit bei gleichzeitiger Platzersparnis zu ermöglichen. Es werden auch die auftretenden strömungsmechanischen Energieverluste durch Reibung sowie Umlenkungen und Ablösungen der Strömung auf eine mögliche Verbesserung überprüft. Abschließend werden die Ergebnisse der Arbeit ausgewertet und auf die Erfüllung der Zielsetzung kontrolliert

Schlüsselwörter: *Thermodynamik, numerische Strömungsmechanik, Kühlkörper, Parameterstudie*

1. EINLEITUNG

Da Kühlkörper viel Platz beanspruchen und oft an Stellen verbaut sind, welche die Wartung erschweren, ist es oft ein großes Anliegen die Außenmaße des Kühlkörpers so weit wie möglich zu reduzieren. In dieser Thesis wird ein Fall behandelt, bei dem der Kühlkörper auf sehr engem Raum verbaut ist und genau dieser Fall eintritt. Die Wartung ist bei eingebautem Zustand sehr schwierig und es wird nach einer Möglichkeit gesucht, den Kühlkörper so klein wie möglich zu gestalten um Platz für weitere Komponenten zu schaffen und die Wartung zu vereinfachen. Der Kühlkörper soll in seiner Breite und Höhe nicht verändert werden, da die Fassung dafür vorgegeben ist. Allerdings soll er, wenn möglich, in seiner Länge reduziert werden um besagten Platz zu schaffen. Es soll dabei eine bessere Kühlleistung ermöglicht werden.
Weiters ist die Finnenform des bis dato genutzten Kühlkörpers für die Anwendung nicht ideal, da die Kühlkanäle in der praktischen Anwendung oft verschmutzen und es danach zu Einbußen bei der Kühlleistung kommt. Die Reinigung der Kühlkanäle gestaltet sich in eingebautem Zustand als aufwändig und es ist nur wenig Raum verfügbar um die Kanäle mit Spezialwerkzeugen zu reinigen. Dadurch erfolgte die Reinigung oft nur unzureichend. Dies führte in der Vergangenheit bei einigen Fällen zur Überhitzung der auf dem Kühlkörper angebrachten Elektronik und es soll deshalb Ziel dieser Arbeit sein, eine für die Wartung geeignetere Form der Kühlfinnen zu finden.

2. PROBLEM- UND AUFGABENSTELLUNG

Zusammengefasst ist das Ziel dieser Arbeit die Kühlleistung des Kühlkörpers zu verbessern und gleichzeitig dessen Platzbedarf zu reduzieren. Es sollen auch Möglichkeiten zur Verbesserung der Wartbarkeit und der strömungsmechanischen Energieverluste durch Reibung, sowie Umlenkungen und Ablösungen der Strömung bei gleicher oder besserer Kühlleistung gefunden werden. Um dies zu bewerkstelligen wurde das CFD-Tool 6SigmaET verwendet, mit dem die benötigten Simulationen durchgeführt werden.

Es werden folgende Forschungsfragen behandelt:

1) Ist es möglich mit einer anderen Finnengeometrie eine höhere Kühlleistung bei gleichzeitig besserer Wartbarkeit zu erreichen?
2) Ist es möglich die Außenmaße des Kühlkörpers bei besserer Kühlleistung zu reduzieren?
3) Kann durch eine Veränderung der Kühlkörpergeometrie der Druckverlustbeiwert und somit die strömungsmechanischen Energieverluste bei besserer Kühlleistung verringert werden?

3. UNTERSUCHUNG DES KÜHLKÖRPERS

Es wurde als erster Schritt ein Hochleistungskühler des Unternehmens Constellium vom Typ HKH-R300RA10 untersucht, welcher als Basis für die nachfolgenden Untersuchungen diente. Es wurde mittels 6SigmaET ein Modell erstellt und anschließend eine Netzunabhängigkeitsstudie durchgeführt um die geeignete Netzgröße für die darauffolgenden Berechnungen zu finden.

3.1 Kühlkörper

Abbildung 1 zeigt den gesamten Kühlkörper HKH-R300RA10. Dieser Kühlkörper wurde in CATIA V5 konstruiert um ihn anschließend mittels 6SigmaET in verschiedenen Szenarien zu simulieren und die Ergebnisse auszuwerten. Es sind ebenfalls die Stromrichtermodule in vereinfachter Form zu sehen, welche die Verlustleistung als Wärme an den Kühlkörper abgeben.

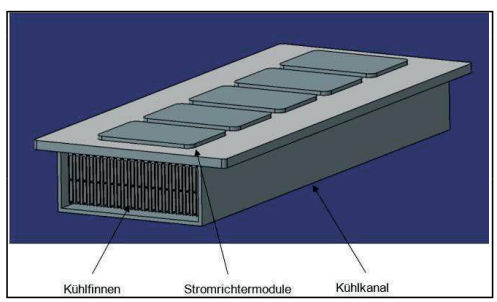


Abb. 1: Kühlkörper, bestückt mit 5 Stromrichtermodulen

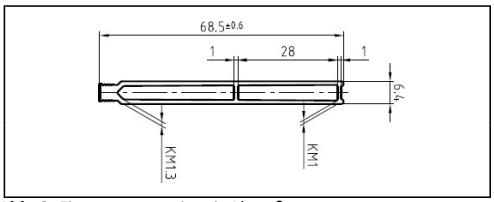


Abb. 2: Finnengeometrie mit Abmaßen

Abbildung 2 zeigt eine einzelne Hohlfinne des Kühlkörpers HKH-R300RA10. Durch die permanente Durchströmung mit Luft und die darin enthaltenen Staub- und Schmutzpartikeln, kommt es oft zu einer Verunreinigung der Hohlkammern.

3.2 Erstellung des Modells

Um die benötigten Daten zur Lösung der Ziele dieser Arbeit zu sammeln, wurden nun mit Hilfe des CFD-Programmes 6SigmaET die notwendigen Simulationen durchgeführt. Dabei wurde das verwendete Stromrichtermodul mittels der CAD-Toolbox des Simulationsprogrammes ausreichend detailgetreu konstruiert um den Belastungsfall realistisch darzustellen. Anschließend wurde die Höhe des Gesamtsystem (Anhang 2) mittels CATIA V5 von $624\ mm$ auf $175\ mm$ reduziert und in verkürzter Form mit nur einem Stromrichtermodul simuliert, da der Simulationsaufwand dadurch viel geringer ausfiel und trotzdem auf das Gesamtsystem rückgeschlossen werden konnte. Die Simulation des ursprünglichen Systems ermöglichte eine Datenbasis auf der dann die nachfolgenden Untersuchungen aufbauten. Nach diesem Schritt war der Weg für eine Optimierung des Kühlkörpers frei und es konnte mit unterschiedlichen Geometrien der Kühlfinnen experimentiert werden.

Das Stromrichtermodul besteht aus insgesamt 8 IGBT´s, welche jeweils eine Verlustleistung von $50\ W$ verursachen, 8 Dioden mit einer jeweiligen Verlustleistung von $25\ W$, 4 NTC- Sensoren welche im Falle eines realen Aufbaus als Messpunkte der Temperatur dienen und einer Grundplatte aus Kupfer. Auf der Grundplatte ist eine Wärmeleitfolie mit einem Wärmeleitwert von $\lambda = 2.5 \frac{\mathrm{W}}{\mathrm{m\,K}}$ angebracht, um den Wärmeübergang in der realen Anwendung effektiver zu gestalten und Lufteinschlüsse, welche die Wärmeübertragung stark negativ beeinflussen würden, zu verhindern.

Um das Modell zu erstellen wurde ein sogenanntes Open-Boundary-Modell erstellt. Der Simulationsraum ist bei diesem Modell nicht geschlossen wie in einer Box, sondern es wird angenommen, dass sich der Simulationsraum in einem unendlichen Raum befindet wobei natürlich die Randbedingungen an den Grenzen des Simulationsraumes gesetzt werden können. Als Material des Kühlkörpers, inklusive des Kühlkanals, ist eine Aluminium- Kupfer-Legierung gewählt, welches dem Material des Originals entspricht.
Die Materialeigenschaften wurden mit einem Wärmeleitwert $\lambda = 164\ W/m\ K$, einer Dichte $\rho = 2790\ kg/m^3$, einer Wärmekapazität $c = 883\ J/kg\ K$ und einer Emissivität von $\varepsilon = 0{,}92$ festgelegt.
Die einzelnen Komponenten wurden anschließend, wie in Abb.11 zu sehen, zusammengefügt und in die gewünschte Position gebracht. Weiters wurde ein Lüfter an die Vorderseite des Modells angebracht und die konstante Durchflussrate wurde mit $V = 200\ m^3/h$ gewählt. Abbildung 3 zeigt das gesamte Simulationsmodell.

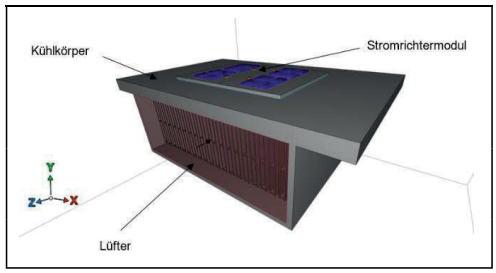


Abb. 3: Gesamtes Simulationsmodell mit erfolgter Zusammensetzung der einzelnen Komponenten

Als Nächstes wurden die Randbedingungen und Eigenschaften des Simulationsraumes festgelegt:
Die Umgebungstemperatur wurde mit $T = 20\ °C$ festgelegt, sowie der Druck an den Grenzen des Simulationsraumes mit $p = 101325\ Pa$ und die Strömungsgeschwindigkeit mit $u = 0\ m/s$, $v = 0\ m/s$ und $w = 0\ m/s$. Es wurde das Medium Luft bei 20°C als inkompressibles, ideales Gas gewählt.
Da es bei turbulenten Strömungen häufig zu oszillierenden Lösungen kommt, die Temperatur aber dabei meist bereits im eingeschwungenen Zustand ist, wurden die Iterationen auf 300 begrenzt. Ob es sich im jeweiligen Fall wirklich so verhält muss allerdings immer nachgeprüft werden. Numerische Experimente haben gezeigt, dass im Durchschnitt unter 300 Iterationen zu einem konvergierten Resultat führen. Die Erdbeschleunigung wurde in Richtung der Y-Achse mit $g_y = -9.81\ m/s^2$ definiert. Um Geometrieränder, beziehungsweise Grenzschichten, richtig zu berechnen und darzustellen, wurden hybride Zellen erlaubt. Dies ermöglicht beim Vernetzen des Modells, dass sich Zellen überschneiden und somit eine genauere Darstellung der Geometrie. Sonnenstrahlung und Wärmestrahlung wurden bei dem Modell nicht berücksichtigt.

3.3 Studie der Netzunabhängigkeit

Um die Strömung des Fluids korrekt darzustellen, musste in essentiellen Bereichen das Netz manuell optimiert werden. Dies geschah mithilfe einer programminternen „Netzkontrolle“, welche erlaubt, an bestimmten Stellen des Modells die Feinheit des Netzes weiter anzupassen.

Die Optimierung wurde mit dem Netz des Moduls gestartet. Dabei wurde das in das Gesamtmodell eingefügte Stromrichtermodul mit unterschiedlichen Vernetzungsgrößen in Y-Richtung berechnet.
Die Einstellung erfolgte über die Maximalgröße $y - Max$ der einzelnen Zellen. In X- und Z-Richtung wurde die Maximalgröße bei $2\ mm$ belassen, da sich bei Verkleinerung dieser Werte nichts am Ergebnis änderte. An einem Messpunkt wurde die Temperaturveränderung in Abhängigkeit der Netzgröße ermittelt. Abbildung 4 zeigt die Position dieses Messpunktes.

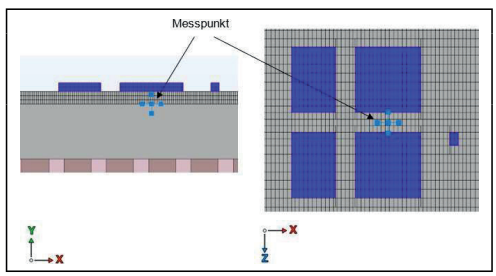


Abb. 4: Position des Messpunktes

Das Modell wurde nun mit unterschiedlichen Konfigurationen berechnet und anschließend wurde mittels eines Vergleichs die beste Konfiguration ausgewählt. Kriterium war, dass sich die Temperatur um nicht mehr als $\Delta T = \pm 0.2\ °C$ ändert.
Da sich ab einer Netzgröße in Y-Richtung von $1\ mm$ nur mehr kleine Schwankungen ergaben, wurde diese Größe für das Modul gewählt

Der Kühlkanal war der nächste Bereich bei dem die Größe des Netzes große Auswirkungen auf das Ergebnis hatte. Im Kanal fand der größte Wärmeübergang statt und um das Strömungsverhalten der Luft ausreichend genau darzustellen, ist eine ausreichende Auflösung notwendig. Gerade an den Grenzschichten kann es durch ein zu grobes Netz zu Fehlern kommen.
Da in Richtung der X-Achse die Abstände zwischen den Kühlfinnen sehr klein sind, wurde bei der Studie der Netzunabhängigkeit der Schwerpunkt auf die Auflösung in Richtung der X-Achse gelegt.
Es konnte ermittelt werden, dass sich die Temperatur am gewählten Messpunkt ab einer Größe von $0.25\ mm$ nicht mehr stark ändert und in der gewählten Toleranz von $\Delta T = \pm 0.2\ °C$ lag.

3.4 Ergebnisse des originalen Kühlkörpers

Das Netz wurde angepasst und dadurch ein Ergebnis ermittelt. Die Ergebnisse wurden nun als Basis für die nachfolgende Variation diverser Parameter zur Optimierung des Kühlkörpers verwendet. Es wurden die Temperatur an den
Messpunkten 1-4, die Strömungsgeschwindigkeit am Eingang, sowie der Druckverlust von Eingang zu Ausgang des Kühlkanals ermittelt.
Abbildung 5 zeigt die Position der Messpunkte, sowie die Oberflächentemperatur des Kühlkörpers.

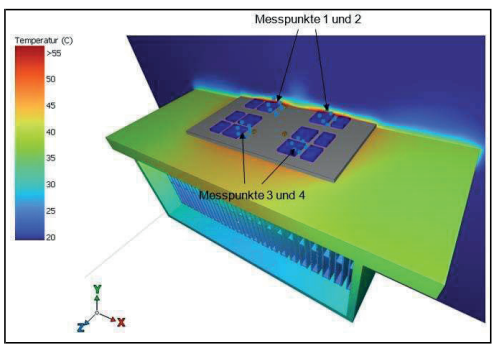


Abb. 5: Oberflächentemperatur des Kühlkörpers und Temperatur auf der Lösungsebene

Tabelle 1 zeigt die wichtigsten Ergebnisse der Simulation. Die Messpunkte 2 und 4 wurden Aufgrund der Symmetrie nicht angeführt.

Messpunkt 1 [°C]	55,5
Messpunkt 3 [°C]	54,1
Gemittelte Strömungsgeschwindigkeit im Querschnitt am Eingang v [m/s]	3,89
Druckverlust Δp_T [Pa]	51

Tab. 1: Ergebnisse, originaler Kühlkörper

4. PARAMETERSTUDIE

Nun sollte mittels einer Parameterstudie festgestellt werden, ob mit einer Veränderung der Geometrie des
Kühlkörpers ähnliche oder bessere Ergebnisse erzielt werden können.

4.1 Ermittlung der Parameter

Es gibt viele Faktoren welche die Temperaturverteilung und den Wärmetransport beeinflussen. Deshalb beschränkt sich die Wahl der Parameter auf die Geometrie des Kühlkörpers und dessen Kühlfinnen.
Als Geometrie der alternativen Kühlfinnen wurde eine simple, quaderförmige Form gewählt. Diese ermöglicht eine gute Sichtbarkeit der Auswirkungen der Parameter sowie eine Verbesserung der Wartbarkeit durch das Wegfallen der hohlen Form der Finnen.
Abbildung 6 zeigt die Form der Finnen.

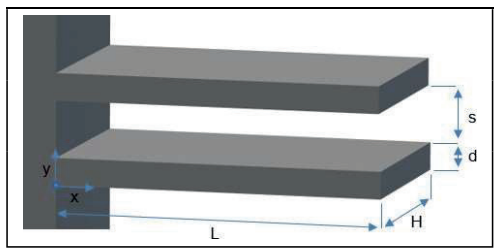


Abb. 6: Quaderförmige Kühlfinne

Die an den Kühlkörper angebrachten Kühlfinnen dienen hauptsächlich zur Vergrößerung der Oberfläche des Kühlkörpers um die Wärmeübertragung an das umgebende Fluid (in diesem Fall Luft) zu verbessern. Die Wahl der Parameter wurde deshalb auf die geometrischen Einflussfaktoren begrenzt.
Zusammenfassend wurde festgestellt, dass es mehrere geometrische Parameter gibt, welche sich auf den Wärmetransport auswirken. Diese sind: die Länge der Kühlrippe L, die Höhe der Kühlrippe H, die Dicke der Kühlrippe d, die Anzahl der Kühlrippen K und der Abstand zwischen den Kühlrippen s. Diese 5 Parameter standen zur Auswahl um sie auf die Auswirkung auf die Kühlleistung zu überprüfen.

4.2. Variation der Parameter d, K, L und s

Folgend wurden die Dicke der Kühlfinnen d, die Anzahl der Kühlfinnen K, der Abstand der Kühlfinnen s, sowie die Finnenlänge L variiert und ermittelt, wie sich die Parameter auf die Kühlleistung auswirkten. Die Höhe der Kühlrippen H wurde identisch der originalen Konfiguration gewählt und nicht verändert. Die Dicke wurde vorerst mit $d = 1\ mm$, $d = 2\ mm$ und $d = 3\ mm$ gewählt, um zu sehen, ob es nötig war, auch mit einer größeren Dicke zu simulieren oder ob sich hier bereits eine Tendenz Richtung Optimum erkennen ließ. Bei der Variation von L wurde auch der Kühlkanal verkleinert, aber die Abstände der äußeren Finnen zum Kanal belassen, wie es auch dem realen Aufbau entsprechen würde.
Es wurde die Entwicklung der Temperatur an den Messpunkten, sowie der Druckverlustbeiwert im Verhältnis zu den Parametern ermittelt.
Der Druckverlustbeiwert ζ berechnet sich mit

$$\zeta = \frac{\Delta p_T}{\frac{\rho}{2} v^2} \qquad (1)$$

wobei Δp_T die Differenz des Druckabfalls des Totaldrucks zwischen Ein- und Ausgang ist, ρ die Dichte der Luft und v die gemittelte Geschwindigkeit am Eingangsquerschnitt [16].

5. ERGEBNISSE

Anschließend wurden die Ergebnisse der Parametervariation in Hinsicht auf die Temperaturentwicklung an den Messpunkten ausgewertet.
Abbildung 7 zeigt die Temperaturverläufe am Messpunkt 1 im Verhältnis zu der Finnenlänge L und des Abstandes zwischen den Kühlrippen s. Es ist zu erkennen, dass die Konfigurationen mit der Länge $L = 38{,}5mm$ die besten Ergebnisse liefert. Ab einem Abstand von $s = 0{,}7mm$ verschlechtert sich die Kühlleistung und die Temperatur steigt.

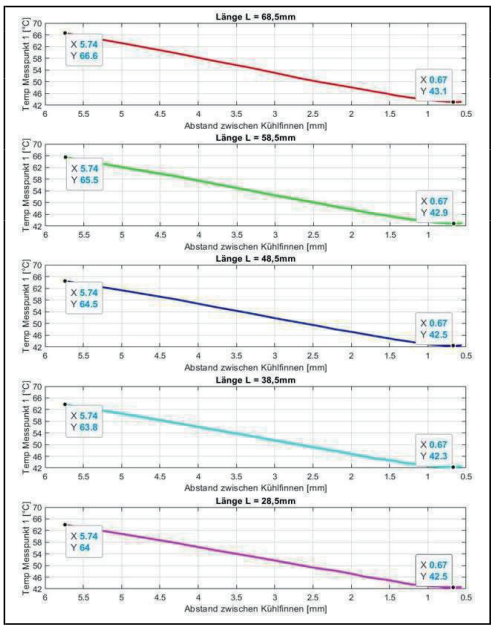


Abb. 7: Temperaturverlauf in Abhängigkeit der Länge L und dem Abstand zwischen den Kühlrippen s bei d = 1 mm

Abbildung 8 zeigt eine dreidimensionale Darstellung der Werte.

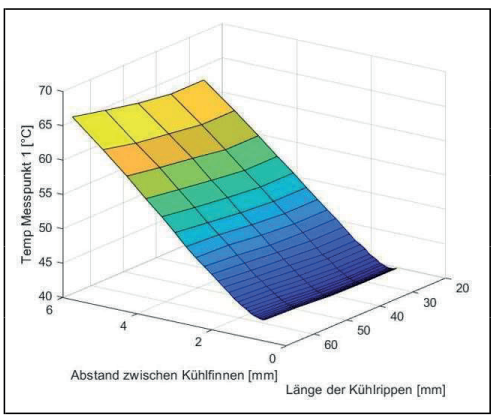


Abb. 8: Dreidimensionale Darstellung der Ergebnisse bei d = 1 mm

Da der ursprüngliche Kühlkörper ein sehr gutes Strömungsverhalten und dadurch einen geringen Totaldruckverlust von $\Delta p_T = 51\ Pa$ und einer auf den Querschnitt gemittelten Geschwindigkeit am Eingang von $v = 3.89\ \frac{m}{s}$ aufwies, sollte auch der Druckverlust sowie die Eingangsgeschwindigkeit bei den alternativen Geometrien untersucht werden.
In Folge dessen wurde der Druckverlustbeiwert ζ für die simulierten Konfigurationen errechnet.
Um den Druckabfall zwischen Ein- und Ausgang zu ermitteln, wurden 2 Lösungsebenen in das Modell eingefügt und anschließend der Durchschnittsdruck ausgelesen. Es wurde auch die durchschnittliche Strömungsgeschwindigkeit v am Eingang ausgelesen. Durch die Berechnung mit dem inkompressiblen, idealen Gas „Luft bei 20°C“ war die durchschnittliche Geschwindigkeit am Eingang genauso groß wie am Ausgang.

Der Druckverlustbeiwert war bei dem ursprünglichen Kühlkörper bei $\zeta = 5.62$ mit einer Temperatur am Messpunkt 1 von $T = 55\ °C$. Nun sollte ermittelt werden, wie sich die anderen Konfigurationen diesbezüglich verhielten.

Die gemittelte Strömungsgeschwindigkeit v am Eingang in Abhängigkeit der durchströmten Fläche A und somit der Finnenlänge L, stellten sich wie in Tabelle 2 zu sehen ein:

Finnenlänge L [mm]	68,5	58,5	48,5	38,5	28,5
Strömungsgeschw. am Eingang v [m/s]	3.89	4.51	5.39	6.7	8.81

Tab. 2: Auf den Querschnitt gemittelte Geschwindigkeit v bei Variation der Finnenlänge L

Abbildung 9 zeigt den Druckverlustbeiwert ζ in Relation zu dem Abstand der Kühlfinnen s bei einer Finnendicke von $d = 1\ mm$.

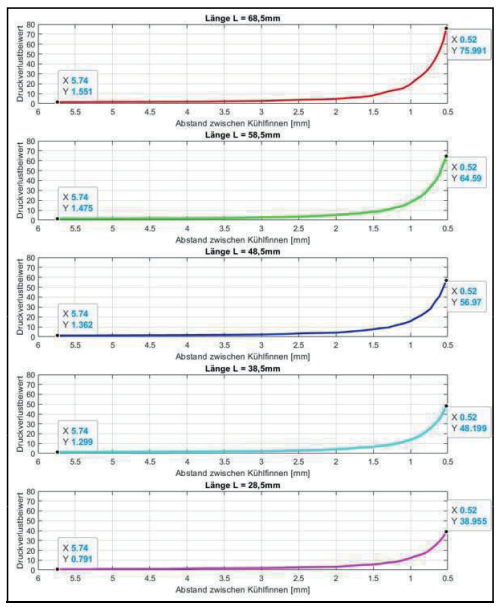


Abb. 9: Druckverlustbeiwert im Verhältnis zum Abstand der Kühlfinnen bei einer Finnendicke von d = 1 mm

6. ZUSAMMENFASSUNG UND AUSBLICK

6.1. Zusammenfassung

Ziel der vorliegenden Arbeit war es, eine alternative Finnengeometrie für einen Kühlkörper zu finden, welche eine einfachere Wartbarkeit des Kühlkörpers ermöglicht. Weiters sollte eine Reduktion der Außenmaße des Kühlkörpers realisiert und der Energieverlust durch Reibung, sowie Umlenkungen und Ablösungen der Strömung verringert werden. Es sollte bei allen Punkten gleichzeitig eine bessere Kühlleistung des Kühlkörpers erfolgen.

Zu diesem Zweck wurde als erster Schritt eine Erläuterung und Visualisierung der Problemstellung bezüglich der Geometrie des Kühlkörpers und eine detaillierte Beschreibung der Zielsetzung dieser Arbeit erstellt. Danach wurde mittels eines 3D-Zeichenprogrammes die ursprüngliche Kühlkörpergeometrie konstruiert und in das CFD-Tool 6SigmaET importiert um anschließend eine erste Simulation zu bewerkstelligen. Um Simulationszeit zu sparen, wurde der Kühlkörper gekürzt und anschließend mit einem Stromrichtermodul bestückt, welches eine Simulation eines realistischen Belastungsfalles ermöglichte. Es wurden die Randbedingungen festgelegt und anschließend eine Netzstudie durchgeführt, um die Simulationsergebnisse unabhängig von der Netzgröße zu machen. Danach wurden Messpunkte definiert, um die Kühlleistung des Kühlkörpers zu messen. Nach diesen Schritten wurden die Temperaturen an den Messpunkten, sowie die über den Querschnitt gemittelte Strömungsgeschwindigkeit am Eingang und der Totaldruckverlust zwischen Ein- und Ausgang ausgelesen. Diese Werte dienten als Ausgangsbasis zur Optimierung des Kühlkörpers.

Als nächster Schritt wurde eine Parameterstudie durchgeführt, bei der eine alternative, quaderförmige Finnengeometrie eingesetzt wurde. Es wurden anschließend die einflussgebenden Parameter der Geometrie des Kühlkörpers ermittelt und darunter 4 Größen ausgewählt, welche anschließend variiert wurden. Diese waren die Finnenlänge, die Finnendicke, der Finnenabstand sowie die Anzahl der Kühlfinnen. Nach einer erneuten Netzstudie wurden die Simulationen der Parameterstudie durchgeführt und wieder die Temperaturen an den Messpunkten, die gemittelte Strömungsgeschwindigkeit am Eingang sowie der Totaldruckverlust zwischen Ein- und Ausgang ausgelesen. Auf Basis der gewonnenen Daten erfolgte eine Darstellung der Temperatur am aussagekräftigsten Messpunkt im Verhältnis zum Finnenabstand. Dies ermöglichte eine Ermittlung der besten Variationen der Kühlkörpergeometrie in Bezug auf die Kühlleistung des Kühlkörpers, sowie der Konfigurationen, welche eine Reduzierung der Außenmaße bei besserer Kühlleistung ermöglichten. Danach wurde der Druckverlustbeiwert für jede Konfiguration berechnet und wieder im Verhältnis zum Finnenabstand dargestellt. Es konnten nun Variationen der Kühlkörpergeometrie ausgemacht werden, welche eine Verringerung des Energieverlustes bei höherer Kühlleistung ermöglichten. Anschließend wurden die Kühlkörpergeometrien, welche die Anforderungen am besten erfüllten, aus den vorliegenden Ergebnissen ermittelt und näher analysiert.

6.2. Fazit und Ausblick:

Im Rahmen der vorliegenden Arbeit konnte die ursprüngliche Hohlfinnengeometrie des Kühlkörpers für den gegebenen Belastungsfall verbessert werden. Die Ziele dieser Arbeit waren eine bessere Wartbarkeit des Kühlkörpers durch eine alternative Finnengeometrie, eine Reduktion der Außenmaße des Kühlkörpers sowie eine Verringerung des Druckverlustbeiwerts und damit der strömungsmechanischen Energieverluste. Diese Punkte sollten alle mit einer Verbesserung der Kühlleistung einher gehen.

Die Ergebnisse der Simulation der ursprünglichen Geometrie des Hochleistungskühlers zeigten bereits ein sehr gutes Strömungsverhalten mit einem geringen Energieverlust und niedrigen Temperaturen an den definierten Messpunkten. Diese Werte dienten als Ausgangspunkt für die darauffolgende Parameterstudie bei der die definierten Parameter der alternativen Finnengeometrie variiert wurden. Die alternative, quaderförmige Finnengeometrie erfüllte zudem die Anforderung nach einer verbesserten Wartbarkeit, da diese nicht mehr hohl gestaltet war. Anhand der ermittelten Temperaturwerte der alternativen Variationen der Kühlkörpergeometrie konnte die Kühlleistung des Kühlkörpers gesteigert werden. Es war dabei zu beobachten, dass die Temperatur an den Messpunkten bei Erhöhung der Anzahl der Kühlfinnen mit einhergehender Reduktion des Finnenabstandes, bis zu einem gewissen Punkt sinkt und anschließend wieder ansteigt. Dies wurde auf die Verdrängung der bewegten Luft durch Grenzschichtbildung zurückgeführt. Die Luft strömte in Folge dessen, nicht optimal genutzt, außerhalb der Kühlfinnen. Es konnte aber bei vielen Konfigurationen eine weitaus niedrigere Temperatur an den Messpunkten erreicht werden als bei der Originalgeometrie. Außerdem war ersichtlich, dass eine Verringerung der Außenmaße des Kühlkörpers mittels einer Verringerung der Finnenlänge bei besserer Kühlleistung des Kühlkörpers erreicht werden konnte. Es wurde anschließend der Druckverlustbeiwert für alle Variationen der Kühlkörpergeometrie berechnet und somit auf das Ausmaß der strömungsmechanischen Energieverluste rückgeschlossen. Trotz des durchschnittlich höheren Druckverlustes der Geometrien bei verkürzter Finnenlänge und somit einer kleineren Querschnittsfläche des Kühlkanals, war der Druckverlustbeiwert aufgrund der höheren auf den Querschnitt gemittelten Strömungsgeschwindigkeit bei einigen Konfigurationen kleiner. Die Energieverluste konnten bei mehreren alternativen Kühlkörpergeometrien verringert werden wodurch eine weitere Anforderung dieser Arbeit erfüllt wurde. Anschließend wurden jene Geometrien aus den Ergebnissen gesucht, welche die jeweiligen einzelnen Anforderungen nach der Reduktion der Außenmaße, des niedrigeren Druckverlustbeiwertes und der erhöhten Kühlleistung am besten erfüllten. Hierbei erfüllte eine der ausgewählten Geometrien alle Anforderungen, siehe Tabelle 8. Die Anforderung der alternativen Finnengeometrie und somit der besseren Wartbarkeit, wurden von jeder alternativen Geometrie erfüllt.

Bei der ausgiebigen Beschäftigung mit der Aufgabenstellung, sowie der Thermodynamik und der numerischen Strömungsmechanik wurden auch andere Problemstellungen und Möglichkeiten zur Weiterführung der Arbeit ersichtlich. Zu diesen zählen zum Beispiel die Variation des Volumenstromes, der Außentemperatur sowie einer weiteren Variation der Finnengeometrie. Da die alternativen Kühlkörpergeometrien eine sehr geringe Finnendicke aufwiesen, wäre eine mechanische Untersuchung auf eine mögliche Entstehung von Schwingungen durch die strömende Luft eine sinnvolle Ergänzung der Arbeit. Auch die Untersuchung des gesamten Systems mit der vollen thermischen Belastung der fünf Stromrichtermodule wäre für die schlussendliche Auswertung der Funktionalität des Kühlkörpers von Interesse.

7. LITERATURVERZEICHNIS

Stephan P., Kabelac S., Kind M., Mewes D., Schaber K., Wetzel T. (2019). VDI- Wärmeatlas. Fachlicher Träger VDI-Gesellschaft Verfahrenstechnik und Ch [1]

Ferziger J. H., Perić M. (2008). Numerische Strömungsmechanik- Springer- Verlag Berlin Heidelberg [2]

Lecheler S. (2011). Numerische Strömungsberechnung, 2. Auflage. Vieweg+ Teubner [3]

Böckh P., Wetzel T. (2009). Wärmeübertragung: Grundlagen und Praxis. Springer Berlin Heidelberg [4]

Herwig H., Moschallski A. (2009). Wärmeübertragung: Physikalische Grundlagen- Illustrierende Beispiele- Übungsaufgaben mit Musterlösungen. Vieweg [5]

Ghaib K. (2019). Einführung in die numerische Strömungsmechanik. Springer Fachmedien Wiesbaden Springer Vieweg [6]

Bschorer S., Buck T. (2018). Technische Strömungslehre- Lehr- und Übungsbuch. Springer Vieweg [7]

Noll H. B. (1993). Numerische Strömungsmechanik- Grundlagen. Springer Verlag Berlin Heidelberg [8]

Olshausen H. G. (1997). VDI- Lexikon Bauingenieurwesen. Springer- Verlag Berlin Heidelberg [9]

Laurien E., Oertel H. (2011). Numerische Strömungsmechanik. Vieweg+ Teubner [10]

Oertel H, Böhle M. (2010). Übungsbuch Strömungsmechanik- Grundlagen, Grundgleichungen, Analytische und Numerische Lösungsmethoden, Softwarebeispiele. Vieweg+ Teubner Verlag [11]

Zahirovic S. (2002). Anwendung der Methoder der Finiten-Volumen auf transsonische Strömungsprobleme in thermischen Turbomaschinen. TU Wien. [12]

Ferziger J. H., Perić M. (2008). Numerische Strömungsmechanik Springer Verlag Berlin Heidelberg [13]

Gheorghiu V. (2011). Technische Thermodynamik und Strömungslehre. HAW Hamburg [14]

Zierep J. (2021). Principles of fluid dynamics- fundamentals, statics and dynamics of fluids. Springer [15]

Shivamoggi B. K. (2022). Introduction to Theoretical and Mathematical Fluid Dynamics. Wiley [16]

Siekmann H. E., Thamsen P. U. (2009). Strömungslehre für den Maschinenbau- Technik und Beispiele. Springer-Verlag Berlin Heidelberg [17]

Wagner W. (2004). Wärmeübertragung. 6. Auflage. Vogel Buchverlag, Kamprath Reihe [18]

Lee S., (1995). Optimum design and selection of heat sinks. IEEE [19]

Chanson H. (2009), Applied Hydrodynamics: An Introduction to Ideal and Real Fluid Flows, CRC Press Taylor & Francis Group [20]



Thomas Thalhammer, B.Sc

FH Wiener Neustadt, 98927@fhwn.ac.at

2019-2021 Projektingenieur Voith Austria GmbH
2019-2023 Masterstudiengang Mechatronik Fachhochschule Wiener Neustadt
2016-2019 Bachelorstudiengang Mechatronik|Mikrosystemtechnik Fachhochschule Wiener Neustadt

ANALYSE EINES KI ALGORITHMUS ZUR ERKENNUNG VON FAHRSPUREN

Marjan Torkamanjahangiri

Kurzfassung: *In dieser Arbeit werden neuronale Netze, die zum Lernen verwendeten Methoden und die Implementierungsschritte von LaneAF, einem KI Algorithmus zur Erkennung von Fahrspüren, diskutiert. LaneAF verwendet zur Spurerkennung binäre Segmentierungsmasken und Affinitätsfelder. Die KI Erkennung von vertikalen und horizontalen Affinitätsfeldern ermöglicht die Segmentierung von Spurbegrenzungslinien. Es werden zwei bekannte neuronale Netze, DLA-34 und ERFNET untersucht und mit dem Tusimple Datensatz trainiert. sodass ein Vergleich und eine Bewertung der Ausgaben des Netzwerks möglich sind. Wie zu erwarten, steigt die mit zunehmender Anzahl an Schichten und der Komplexität des Netzwerkes der Trainingsaufwand. ERFNET ist einfacher, besitzt eine kurze Trainingszeit, liefert aber schwächere Ergebnisse als DLA-34.*

Schlüsselwörter: *Neuronale Netze, Fahrspurerkennung, Faltungsnetzwerk, Aggregation.*

1. EINLEITUNG UND AUFGABENSTELLUNG

Verkehrssicherheit ist heutzutage zu einem der wichtigsten Faktoren in der Entwicklung von Mobilität geworden. Die meisten Autounfälle werden durch fahrerbezogene Ursachen wie Ablenkung, Müdigkeit und schlechte Fahrleistung verursacht. Eine Verbesserung sind intelligente Autos, welche die menschlichen Faktoren teilweise eliminieren können. Intelligente Fahrzeuge unterstützen die Menschen auf der Grundlage von Echtzeit Verkehrsinformationen. Einer der Faktoren, der sich direkt auf die Umsetzung des Fahrverhaltens auswirkt, ist die Fahrspurerkennung, die einen Meilenstein auf dem Weg der intelligenten Fahrzeugentwicklung darstellt. Anhand der Leitlinien ist es möglich, die ideale, korrekte Position für das smarte Fahrzeug zu bestimmen. Diese Funktion trägt wesentlich dazu bei, die Fahreffizienz und Sicherheit beim automatischen Fahren zu verbessern. Die Methoden zur Spurerkennung werden laufend verbessert um auch bei schwierigen Verhältnissen (Sicht, Regen, Verkehr) gute Ergebnisse mit einem sehr geringen Fehleranteil zu erhalten.

Ziel ist die Entwicklung eines neuronalen Netzes mit geringerer Komplexität, aber hoher Leistung. 2016 wurde ein effizientes Linienerkennungsverfahren vorgestellt, das eine Aufnahme aus der Fahrerperspektive in die Vogelperspektive transformiert und anschließend analysiert. Es wird eine Hough-Transformationstechnik verwendet um eine gekrümmte Fahrspur zu erkennen. Die Parameter der Fahrspur werden unter Verwendung eines Ansatzes der kleinsten Fehlerquadrate geschätzt [Dorj 2016].

2019 wurde der Line Detection Algorithm für komplexen Straßenverhältnisse in dynamischen Umgebungen veröffentlicht. Auch hier wird das Bild in der Draufsicht analysiert, wofür eine Region-of-Interest-Extraktion (ROI) und eine inverse perspektivische Transformation erforderlich sind. Dann wird auf der Grundlage des RANSAC Algorithmus (Random Sample Consensus) die Krümmung der erkannten Fahrspuren angepasst. Unter Verwendung von Straßenfahrvideos unter komplexen Straßenbedingungen und des Tusimple-Datensatzes [Tusimple 2023] wird die Leistungsfähigkeit des Spurerkennungsalgorithmus demonstriert [Jingwei 2019].

2018 wurde ein räumlich-zeitliches, auf Deep Learning basierendes Verfahren zur Erkennung von Fahrspuren vorgeschlagen, das Fahrspurbegrenzungen unter komplexen Wetterbedingungen und Verkehrsszenarien in Echtzeit erkennen kann. Der Algorithmus besteht aus drei Teilen: i) inverse perspektivische Transformation und Fahrspurbegrenzungspositionsschätzung unter Verwendung von räumlichen und zeitlichen Einschränkungen ii) Grenztypklassifizierung und Positionsregression auf Basis von Convolutional Neural Networks (CNN), iii) Optimierung und Fahrspuranpassung [Y. Huang und S. Chen, 2018].

Die Herausforderungen bei der Erkennung der Fahrspur umfassen i) geringe Genauigkeit unter verschiedenen Bedingungen und Szenarien, ii) niedrige Bildverarbeitungsgeschwindigkeit aufgrund hoher Rechenkomplexität, iii) Notwendigkeit eines leistungsfähigen Systems für Training und Verarbeitung.

2. STRUKTUREN NEURONALER NETZWERKE

Zur Analyse von Bilddaten werden meist Faltungsnetzwerke (CNN – convolutional neural networks) eingesetzt [Alzubaidi 2021], die in der einfachsten, klassischen Konfiguration als serielles Netzwerk aufgebaut sind. Dabei werden die Schichten schrittweise durchlaufen, die Ausgänge einer Schicht bilden die Eingänge der nächsten Schicht. Es hat sich aber gezeigt, dass die Verwendung von Ausgängen mehrerer Schichten, die als Aggregation bezeichnet wird, deutlich bessere Ergebnisse liefert. Daher werden in dieser Arbeit aggregierte Modelle untersucht. Das DLA-34 [Liu 2020] ist eines der DLA-Modelle (Deep-LayerAggregation) [Yu 2018], die zur Durchführung der Bildklassifizierung entwickelt wurden. Dieses Modell wird in PyTorch implementiert und steht vortrainiert zur Verfügung.

2.1 Deep Layer Aggregation

Bei der Aggregation vieler Schichten (DLA – Deep Layer Aggregation) werden Schichtenausgänge weit entfernt liegender Schichten miteinander verarbeitet. Die Schichten werden in Blöcke gruppiert, die dann selbst nach ihrer Feature-Auflösung in Stufen gruppiert werden. Dabei haben sich zwei Methoden etabliert, Iterative Deep Aggregation (IDA) und Hierarchial Deep Aggregation (HDA). Dazu werden Aggregationsknoten eingeführt, welche Eingänge unterschiedlicher Dimension mit Hilfe von Projektionsverfahren kombinieren. Aggregationsknoten stellen keine Anforderungen an die interne Struktur der darauffolgenden Blöcke.

Iterative Deep Aggregation

Bei IDA beginnt die Aggregation im flachsten, kleinsten Maßstab und führt dann iterativ tiefere, größere Maßstäbe zusammen. Auf diese Weise werden flache Merkmale verfeinert, während sie durch verschiedene Stadien der Aggregation propagiert werden [Yu 2018], siehe Abb. 1

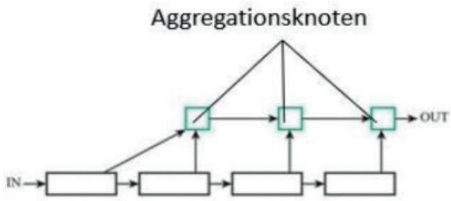


Abb. 1: Iterative Deep Aggregation [Yu 2018]

Hirarchical Deep Aggregation

Zur weiteren Nutzung und Kombination der Merkmalskarten werden hierarchische tiefe Aggregationsblöcke und Stufen in einem Baum zusammengeführt und dabei flachere und tiefere Schichten kombiniert, siehe Abb. 2.

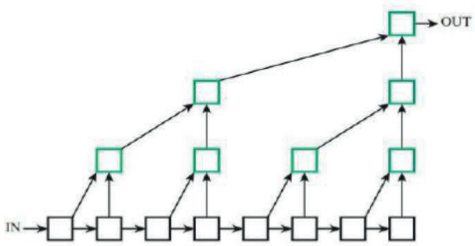


Abb. 2: Hirarchical Deep Aggregation [Yu 2018]

Anstatt nur zwischengeschaltete Aggregationen im Strukturbaum weiterzuleiten kann die Ausgabe eines Aggregationsknotens auch zurück in das Ausgangsnetzwerk (Backbone) geleitet werden und als Eingabe für den nächsten Teilbaum dienen [Yu 2018]. Diese Konfiguration wird als „reentrant Aggregation“ bezeichnet und ist in Abb. 3 gezeigt.

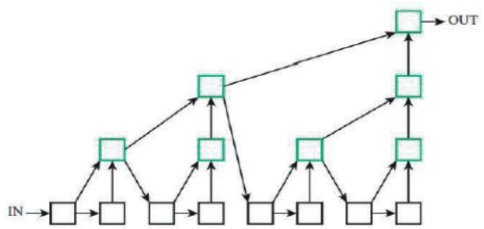


Abb. 3: “Reentrent Aggregation” [Yu 2018]

Dabei kann die Struktur vereinfacht werden, wenn die Anzahl der Aggregationsknoten reduziert wird, siehe Abb. 4.

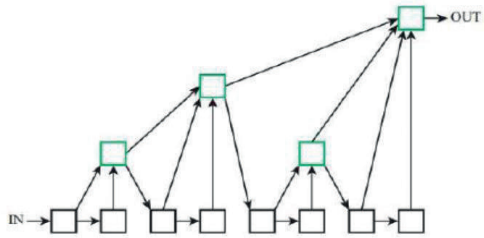


Abb. 4: vereinfachte Konfiguration der „Reentrent Aggregation“ [Yu 2018]

Kombination von HDA und IDA

Die Kombination von IDA und HDA liefert eine leistungsstarke Struktur, bei der die Elemente der IDA durch HDAs mit steigender Komplexität ersetzt werden, siehe Abb. 5.

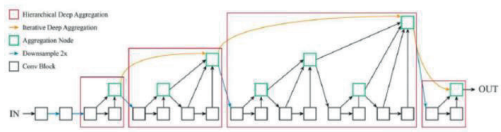


Abb. 5: Kombination von HDA und IDA [Yu 2018]

2.2 ERFNET

ERFNET (Efficient Residual Factorization of Convolutional Networks) ist ein Netzwerk zur effizienten Segmentierung von Bildern das mit einfachen Aggregationsknoten arbeitet. Dabei werden die Schichten so gestapelt, dass ihre Lernleistung und –effizienz optimal genutzt wird. Es besteht aus den zwei Teilen Encoder und Decoder, siehe Abb. 6.

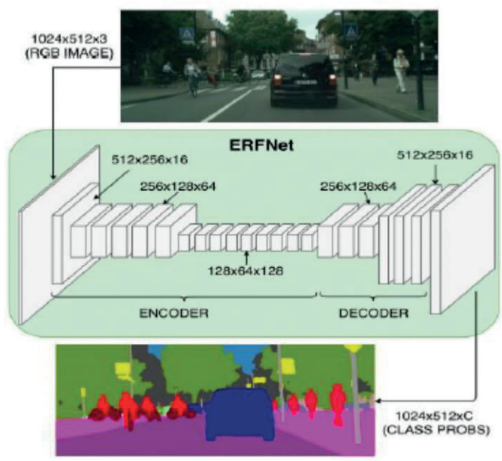


Abb. 6: Struktur des ERFNET Netzwerkes [Romera 2017]

Encoder

Die Schichten von 1 bis 16 der Architektur bilden den Encoder, der aus Restblöcken und Downsampling-Blöcken (Schichten 1, 2 und 8) besteht. „Beim Downsampling werden

einfache CNN-Architekturen verwendet und abstrakte Darstellungen des Eingangsbildes erzeugt [Romera 2017].

Decoder

Das Decoder Segment setzt sich aus den Schichten 17 bis 23 zusammen. ERFNet hat einen kleinen Decoder. Es werden Dekonvolutionsschichten mit Schrittweite 2 verwendet [Romera 2017]. Die einzelnen Schichten von Encoder und Decoder sind aus Elementarblöcken aufgebaut, vgl. Abb. 7.

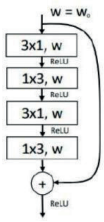


Abb. 7: Detail einer ERFNET Schicht [Romera 2017]

2.3 LaneAF

LaneAF (robust multi-LANE detection with Affinity Fields) ist ein Algorithmus zur effizienten Erkennung von Fahrspuren, der mit Hilfe von neuronalen Netzen (DLA oder ERFNET) aus Eingangsbildern Affinitätsfelder und Binärbilder der Fahrspurbegrenzung generiert.

Dann können diese Affinitätsfelder zusammen mit den Binärmasken verwendet werden, um Bahnpixel in einem Nachbearbeitungsschritt mit Hilfe traditioneller Programmierung bzw. Verarbeitungsschritte zu gruppieren. Dieser Ansatz ermöglicht es, eine variable Anzahl von Fahrspuren zu erkennen, ohne eine feste oder maximale Anzahl von Fahrspuren anzunehmen [Abualsaud 2021].

Für die Erkennung von Fahrspuren wird eine binäre Klassifizierung verwendet. Um verschiedene Spuren zu trennen, wird ein Clustering Schema vorgeschlagen, das auf Affinitätsfeldern basiert. Affinitätsfelder wurden ursprünglich in [Bertasius 2015] zur 2D-Posenschätzung mehrerer Personen eingeführt. Sie bestehen aus Einheitsvektoren, die Ort und Orientierung codieren. In [Abualsaud 2021] werden das horizontale Affinitätsfeld (HAF) und das vertikale Affinitätsfeld (VAF) eingeführt, um Spuren eindeutig zu identifizieren und zu segmentieren. Ein Einheitsvektor im VAF kodiert die Richtung, in der sich der nächste Satz von Spurpixeln in der Bildzeile darüber befindet.

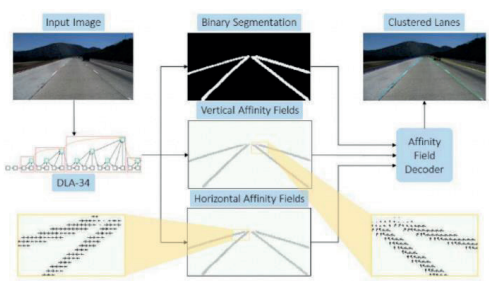


Abb. 8: Deep Layer Aggregation. Schematische Darstellung der Abläufe und Verfahrensschritte [Abualsaud 2021]

Andererseits zeigt ein Einheitsvektor im HAF in Richtung der Mitte der Fahrspur der aktuellen Reihe, wodurch Fahrspuren beliebiger Breite gruppiert werden. Beide Affinitätsfelder können in Verbindung mit der binären Segmentierung verwendet werden, um Vordergrundpixel in einem Nachverarbeitungsschritt zu Fahrspuren zu gruppieren. Eine Schematische Darstellung der in [Abualsaud 2021] verwendeten Verfahrensschritte ist in Abb. 8 gezeigt.

Verlustfunktionen

Bei der Arbeit an einem Maschine Learning oder Deep Learning Problem, werden Verlust-/Kostenfunktionen verwendet, um das Modell während des Trainings zu optimieren [Abualsaud 2021]. Das Ziel ist, eine passend definierte Verlustfunktion zu minimieren, denn je geringer der Verlust, desto besser das Modell. Kreuz-Entropie-Verluste bilden eine sehr wichtige Klasse von Kostenfunktionen, die oft bei Klassifizierungsaufgaben eingesetzt werden. Für den binären Segmentierungszweig wird der gewichtete binäre Kreuz-Entropie-Verlust (Standardverlust) verwendet. Die vom Modell berechneten Ausgangsgrößen werden zum Kreuz-Entropie-Verlust umgerechnet

$$L_{BCE} = -\frac{1}{N}\sum_{i=1..N} w \cdot (t_i \cdot \log(o_i) + (1 - t_i) \cdot \log(1 - o_i))$$

wobei t_i der Zielwert für das Pixel i und o_i die Ausgabe des Netzes darstellt. Das Gewicht w wird verwendet, um Vordergrundpixel stärker zu gewichten. Um den im Bezug auf Vorder- und Hintergrundfläche unausgeglichenen Datensatz weiter zu normieren, wird für den Segmentierungszweig eine zusätzliche Verlustfunktion L_{IoU} (Intersection over Union Loss) verwendet, die in Abb. 9 graphisch dargestellt ist [Lindström 2019].

$$IoU = \frac{Area\ of\ Intersection\ (\text{Overlap})}{Area\ of\ Union} = \frac{t \cap O}{t \cup 0}$$



Abb. 9: Verlustfunktion L_{IoU} (Intersection over Union Loss) mit Zielflächen (Target t) und Netzausgang-Flächen (Output o)

Mathematisch kann die graphische Darstellung von Abb. 7 durch

$$L_{IoU} = \frac{\sum_i t_i \cdot o_i}{\sum_i t_i + o_i - t_i \cdot o_i}$$

approximiert werden. In dieser Beschreibung ist die Verlustfunktion auch differenzierbar, eine wichtige Eigenschaft bei der Optimierung des Netzwerkes.

Intersection over Union, bzw. der Jaccard Koeffizient, ist eine einfache Bewertungsmetrik, die verwendet wird, um die Genauigkeit eines Objektdetektors für einen bestimmten Datensatz zu messen. Jeder Algorithmus, der vorhergesagte Begrenzungsflächen als Ausgabe bereitstellt, kann mit IoU ausgewertet werden.

Für die Affinitätsfeldzweige des Modells wird zusätzlich noch der L_1-Regressionsverlust L_{AF} nur auf die Vordergrundpositionen (foreground fg) sowohl der vertikalen als auch der horizontalen Affinitätsfelder definiert,

$$L_{AF} = \frac{1}{N_{fg}} \sum_{i=1..N_{fg}} \left|t_i^{HAF} - o_i^{HAF}\right| + \left|t_i^{VAF} - o_i^{VAF}\right|$$

Damit ergibt sich der Gesamtverlust L_{tot} des Modells als Summer der Einzelverluste

$$L_{tot} = L_{BCE} + L_{IoU} + L_{AF}.$$

Zur Bewertung der Genauigkeit der Testdaten wird die Genauigkeit

$$Accuracy = \frac{N_{pred}}{N_{gt}}$$

ausgewertet, wobei N_{pred} der Anzahl der korrekt vorgesagten Fahrspurpunkte entspricht, und N_{gt} die Anzahl der Ground Truth Fahrspurpunkte bezeichnet.

3. PRAKTISCHE DURCHFÜHRUNG UND ERGEBNISSE

In diesem Abschnitt werden die Ergebnisse des eigenständigen Trainings der Netzwerke mit dem TUsimple Datensatz vorgestellt.

Seit vielen Jahren verwenden KI Entwickler Grafik Prozessoren (GPU- graphical processing unit), um eine große Menge elementarer Berechnungen schnell und vor allem gleichzeitig (parallel) durchzuführen. Dazu wurde die Programmierschnittstelle Cuda (Compute Unified Device Architecture) entwickelt, die unabhängig von der tatsächlich verfügbaren Hardware ist und auch ohne GPU ausgeführt werden kann. Allerdings dauert dann die Berechnung deutlich länger und die CPU ist völlig ausgelastet.

3.1 DLA-34 Netzwerk

Das publizierte DLA-34 Netzwerk wird mit Hilfe des PyTorch Frameworks für maschinelles Lernen von Python aus trainiert. Das Training mit der Hilfe der CPU dauert für eine Epoche etwa einen Tag, und liefert ein PyTorch-Standardmodell welches für die weitere Nutzung direkt verwendet werden kann. In Abb. 10 ist die Reduktion der Verlustfunktionen L_{AF}, L_{IoU}, L_{BCE} und L_{tot} gezeigt.

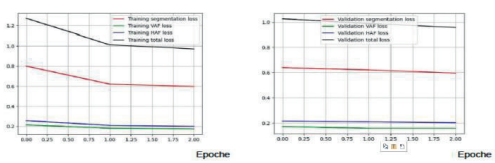


Abb. 10: Training- und Validierungverlust des DLA-34 Netzwerkes

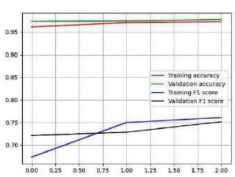


Abb. 11: Genauigkeit (Accuracy) des DLA-34 Netzwerkes

Von den Gesamtdaten, werden, 80% zum Training und 20% zur Validierung verwendet. Die Genauigkeit (Accuracy), siehe Abb. 11 zeigt, dass das verwendete Netzwerk bereits nach 2 Trainingsepochen gut funktioniert.

Um eine Ausgabe des DLA-34 Netzwerks zu erhalten, muss jedes Eingangsfoto eine Größe von 1280x720 Pixel aufweisen. Nach dem Training wird das Netzwerk mit eigenen Daten getestet. Beim Betrachten von typischen Ergebnissen, siehe Abb. 12, ist ersichtlich, dass dieses Netzwerk sehr gut (aber nicht perfekt) funktioniert und die Fahrspuren korrekt erkannt werden. Dabei ist darauf hinzuweisen, dass das Netzwerk nicht mit Fahrszenen aus Europa trainiert wurde, weil der Tusimple Datensatz aus 6408 Bilden von amerikanischen Autobahnen besteht, die bei unterschiedlichen Wetter und Lichtverhältnissen aufgenommen wurden.

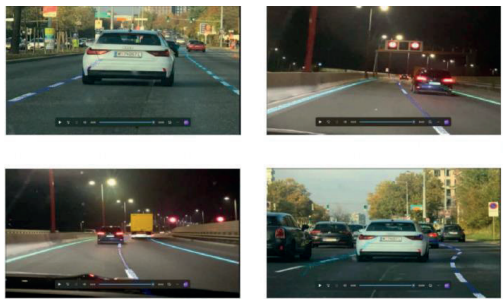


Abb. 12: Straßenszene in Wien mit Fahrspurerkennung durch LaneAF Algorithmus mit DLA -34 Netzwerk

3.1 ERFNET Netzwerk

Zum Vergleich wird LaneAF auch mit der ERFNET Implementierung trainiert. ERFNET ist deutlich einfacher als DLA-34, das Training dauerte pro Epoche nur ca. 2 bis 3 Stunden (CPU-Training). Ein Training für 40 Epochen zeigt, dass sich ab der 20. Epoche kaum mehr Verbesserungen ergeben. Im Vergleich zum DLA-34 Netzwerk (nach 2 Trainingsepochen) sind dabei sind alle Ergebnisse etwas schwächer, siehe Abb. 13-14.

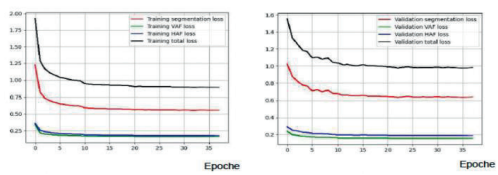


Abb. 13: Training und Validierung Verlust des ERFNET Netzwerkes

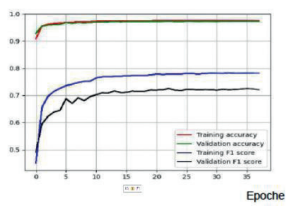


Abb. 14: Genauigkeit (Accuracy) des ERFNET Netzwerkes

Abb. 15 zeigt die Ergebnisse des ERFNET Algorithmus mit einigen Testbildern für typische Fahrsituationen auf österreichischen Autobahnen.

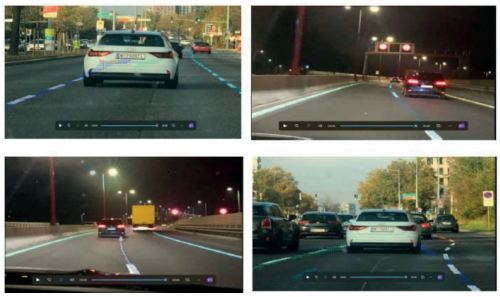


Abb. 15: Straßenszene in Wien mit Fahrspurerkennung durch LaneAF Algorithmus mit ERFNET

Wie das Genauigkeitsdiagramm zeigt, ist die Genauigkeit von DLA-34 höher als jene von ERFNET, was auch aus den visuellen Ergebnissen ersichtlich ist.

3. ZUSAMMENFASSUNG UND AUSBLICK

In dieser Arbeit wird die Fahrspurerkennung mit Hilfe des LaneAF Algorithmus mit zwei unterschiedlichen Bildverarbeitungs- bzw. Segmentierungsnetzwerken untersucht. Die neuronalen Netze unterscheiden sich dabei in ihrer Komplexität, nutzen aber beide das Konzept von Aggregationsknoten, und somit die Zusammenführung von Merkmalsinformationen aus unterschiedlichen Schichten. Die Aggregation ist ein entscheidender Aspekt der Architektur, und mit zunehmender Anzahl der Module wird ihre Konnektivität noch wichtiger. Die Ergebnisse zeigen gute Übereinstimmung mit publizierten Resultaten. Das mit Hilfe des Tusimple Datensatzes trainierte System zeigt eine gute Erkennung von Fahrspuren auf österreichischen Autobahnen, auch wenn diese nicht im Trainingsdatensatz enthalten sind.

7. LITERATURVERZEICHNIS

H. Abualsaud, S. Liu, 2021, LaneAF: Robust Multi-Lane Detection with Affinity Fields, arXiv preprint arXiv:2103.12040

L. Alzubaidi und J. Zhang 2021, concepts, CNN architectures, challenges, applications, future directions, *Journal of Big Data*

G. Bertasius, J. Shi, 2015, A multi-scale bifurcated deep network for top-down contour detection, *Computer Vision and Pattern Recognition,* arXiv:1412.1123v3

B. Dorj und D. J. Lee, 2016, A Precise Lane Detection Algorithm Based on Top View Image Transformation and Least-Square Approaches, Journal of Sensors, vol. 2016, https://doi.org/10.1155/2016/4058093

Y. Huang und S. Chen, 2018, Spatial-Temproal Based Lane Detection Using Deep Learning, *Artificial Intelligence Applications and Innovations*, pp. 143-154

C. Jingwei und S. Chuanxue, 2019, Lane Detection Algorithm for Intelligent Vehicles in Complex Road Conditions and Dynamic Environments, Sensors 2019, 19(14), 3166; https://doi.org/10.3390/s19143166

A. Lindström, E. Okafor, 2019, Deep Neural Networks with Intersection over Union Loss for Binary Image Segmentation, *International Conference on Pattern Recognition Applications and Methods*

L. Liu, L. Haiyu, 2020, Efficient Foreign Object Detection Between PSDs and Metro Doors via Deep Neural Networks, *IEEE Access, vol. 8*, pp. 46723-46734

X. Pan, J. Shi, 2019, Spatial as deep: Spatial CNN for traffic scene understanding, *Computer Vision and Pattern Recognition,* arXiv:1712.06080

E. Romera, R. Arroyo, 2017,ERFNet: Efficient Residual Factorized ConvNet for Real-time Semantic Segmentation, *IEEE Transactions on Intelligent Transportation Systems*

Tusimple 2023, https://github.com/TuSimple/tusimple-benchmark, abgerufen am 22.3.23

F. Yu, D. Wang, 2018, Deep Layer Aggregation, *Conference on Computer Vision and Pattern Recognition (CVPR) 2018,* arXiv:1707.06484v3

AUTOR



BSc Marjan Torkamanjahangiri

Masterstudium Mechatronik - FH Wiener Neustadt, 2700 Wiener Neustadt - 10/2020 – 01/2023, Bachelorstudium Computer Software Ingenieur - Roudehen Azad Universität, Teheran / Iran -10/2006 – 09/2009, Angewandten Studiengang Computer- Software - Roudehen Azad Universität, Teheran / Iran- 10/2002 – 09/2005.

Beruflicher Werdegang:
Software developer – Österreich - (Teilzeit), Business Intelligence – Österreich - (Teilzeit), IT-Expertin - Iran - (Vollzeit)

ANALYSE UND AUTOMATISIERUNG EINES HOCHLASTTRIBOMETERS

Vattappillil B.

Kurzfassung: *In der vorliegenden Arbeit wird ein bestehendes Hochlasttribometer und dessen Komponenten weiterentwickelt. Das untersuchte Tribometer ist ein Kniehebelversuchsstand,d.h. ein Testgerät zur Ermittlung von Verschleiß- und Reibeigenschaften hochbelasteter Gleitlager unter kleinen Bewegungswinkeln. Zum bestehendem System wird eine Antriebssteuerung implementiert, die Sensorik überarbeitet und der Schaltkasten erweitert, um die Bedienbarkeit des Gerätes zu vereinfachen, die Versuchsgenauigkeit zu erhöhen und den Versuchsablauf zu automatisieren. Dabei werden mehrfach neu hinzugefügte Komponenten mit der Steuerung kombiniert und die dazu notwendige Logik für die Steuerung programmiert.*
Ein Mehrkörper-Simulationsmodel auf Basis der bestehenden Konstruktion wird in Simscape Multibody entwickelt, um den Bewegungsablauf im Test genau zu untersuchen.
Um die Verbesserungen zu erproben und den Bewegungsablauf zu verifizieren, werden mit dem modernisierten Teststand zwei Versuche gefahren. Anhand der Messdaten werden in weiterer Folge Auswertealgorithmen programmiert. Im Rahmen dieser Arbeit werden die Messdaten hinsichtlich Verschleiß-, Temperatur- und Kraftentwicklung in MATLAB analysiert und visualisiert. Die Versuche werden hinsichtlich ihrem Verschleißverhalten miteinander verglichen.

Schlüsselwörter: *Versuchsstand, Verschleiß, Reibung, Tribologie, Lager, Mehrkörpersimulation*

1. EINLEITUNG

Der Terminus „Tribologie“ wird nach einer umfassenden Studie über die technologische und wirtschaftliche Bedeutung von Verschleiß, Reibung und Schmierung erstmals 1966 mit der folgenden Definition in England geprägt: „Tribology is the science and technology of interacting surfaces in relative motion and of related subjects and practices“ [Horst und Habig 2020]. Unter einem tribologischem System, oder kurz Tribosystem (Tribos, griechisch reiben) wird im allgemeinem ein aus vier Komponenten bestehendes System verstanden: Grundkörper, Gegenkörper, Zwischenstoff und Umgebungsmedium: der Gegenkörper drückt gegen den Grundkörper mit einer gewissen Belastung und bewegt sich. Ein Zwischenstoff, z.B. ein Schmiermittel, soll die Reibungseigenschaften optimieren. Das Umgebungsmedium oder die Umgebungsbedingungen wirken von außen auf das System ein, z.B. mit erhöhter Temperatur. Die Orte, an denen physikalische Wechselwirkungen zustande kommen, werden nach der Terminologie der Konstruktionslehre als Wirkorte bezeichnet. Die Funktion von einem tribologischen System wird daher über Wirkflächen realisiert. Diese werden durch funktionelle Kräfte und Relativbewegungen tribologisch beansprucht. Verschleiß und Reibung resultieren aus Dissipationseffekten in zeitlich und örtlich stochastisch verteilten Kontakten in Anhängigkeit von Systemstruktur und den Beanspruchungen (Temperatur, Geschwindigkeit, Kräfte, etc.) [Horst und Habig 2020]. Einige der Hauptziele der Tribologie sind die Steigerung der Qualität von Geräten, insbesondere durch Analyse und Optimierung von Kontakteigenschaften von Materialien. Ebenso spielt die Erhöhung der Wirtschaftlichkeit durch Senkung von Reibungsverlusten, durch Reduktion von verschleißbedingtem Material- bzw. Rohstoffeinsatz, sowie durch Vermeidung bzw. Verringerung von Stillstands- und Folgekosten eine wesentliche Rolle: Tribologische Kontakte führen immer unweigerlich zu Energieverlust und in weiterer Folge zu Materialverschleiß, eine Optimierung ist daher immer wünschenswert. So wird z.B. in Lagern eine Minimierung der Reibzahl angestrebt, um Energieverluste zu minimieren. In anderen Systemen, z.B. Verschleißschutzplatten liegt wiederum die Reduzierung des Materialverlustes im Vordergrund. Um derartige Optimierungen wissenschaftlich zu erzielen, werden Tribometer verwendet – Testgeräte mit denen tribologische Kontaktbedingungen nachgestellt werden, um optimale Prozessparameter oder Werkstoffkombinationen für gegebene Anwendungen zu realisieren. Tribometer können von einfachen Modellkontakten (z.B. Kugel gegen Platte) bis hin zu Komponententests reichen (z.B. ganze Lagerbaugruppen), womit auch die Komplexität und der Kosteneinsatz für die Testroutine steigen.

Die Arbeit beschäftigt sich mit einem Kniehebelteststand, welcher von der Firma AC2T research GmbH (AC²T) entwickelt wurde. AC²T ist das Österreichische Exzellenzzentrum für Tribologie, ein Forschungsinstitut welches seinen Standort in Wiener Neustadt hat. Die Firma beschäftigt sich mit allen Teilgebieten der Tribologie, von der Konstruktion, über Werkstoffoptimierungen und Schmierstoffentwicklung bis hin zur numerischen Auslegung von Tribosystemen. Der Kniehebelteststand, bietet die Möglichkeit zur experimentellen Bestimmung von Verschleißeigenschaften bestimmter Gleitlager unter sehr großen Lasten bei gleichzeitig geringen Auslenkungen. Derartige Belastungen kommen z.B. in Kniehebelverschlussvorrichtungen von Spritzgussformen vor. Das Konzept ist folgendes: Auf beiden Seiten eines symmetrischen Testaufbaus, sind Testlager eingebaut, die mit einer vorgegebenen Kraft beansprucht werden. Über zwei Antriebe werden, die eingebauten Lager über einen kleinen Winkel dynamisch radial belastet und bei variierenden Geschwindigkeiten und Belastungen getestet.

2. PROBLEM- UND AUFGABENSTELLUNG

Angesichts aufwändiger Versuchsvorbereitungen und Ungenauigkeiten, die während des Versuchs auftreten (vor allem ein Drift der Belastung), die jedoch erst im Nachhinein bei der Versuchsauswertung sichtbar werden, wurde entschieden den vorhandenen Teststand zu erneuern. Im Rahmen dieser Überarbeitung soll die Bedienbarkeit des Gerätes verbessert und Änderungen vorgenommen werden um stabile Versuche zu garantieren. Der Verschleiß an den Lagern war in der Ausgangssituation nur nach dem Test gemessen werden, hier soll eine zusätzliche online Verschleißmessung ergänzt werden. Zudem sollen Simulationen die Belastung und den Verschleißabtrag validieren bzw. analysieren.

3. AUFBAU UND BESCHREIBUNG DES TRIBOMETERS

Die Hauptfunktion des Tribometers (siehe Abb. 1) ist die Testlager in denen die Testwelle gelagert ist zyklisch hohen Lasten bis 160 kN auszusetzen. Die Be- und Entlastung soll über einen kleinen Schwenkwinkel möglichst rasch erreicht werden. Die Bewegung wird mit einer Kurbelscheibe und einem Hebelmechanismus umgesetzt, womit eine zentral sitzende Tellerfeder zusammengedrückt wird. Die Federkraft entspricht dann der Belastung des Lagers. Der Teststand ist symmetrisch aufgebaut und prüft zwei gleichartige Lager simultan.

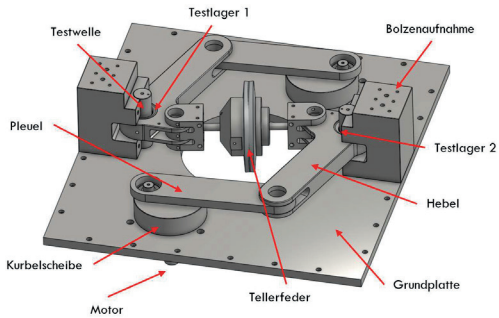


Abb. 1*:* Hauptmaschinenkomponenten des Kniehebelteststandes

Der Hebel (1), siehe Abb. 2 ist mit dem Testlager (2) verbunden, welches auf der Testwelle (3) sitzt. Die Aufgabe des Hebels besteht darin die Kraft des Motors zu verstärken, durch die Länge r_k wird die Kraft F_k erhöht. Das statische Momentengleichgewicht (2D) wird zur Veranschaulichung aufgestellt und anschließend umgeformt:

$$M = \sum_{i=1}^{n} M_i = \sum_{i=1}^{n} F_i \cdot r_i \tag{1}$$

$$\sum M_A = F_F \cdot r_F - F_k \cdot r_k = 0 \tag{2}$$

$$F_F = F_k \cdot \frac{r_k}{r_F} \tag{3}$$

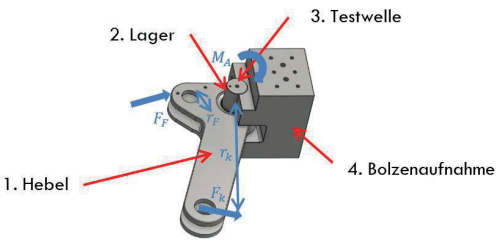


Abb. 2*:* Lasthebel

Damit gelingt es durch mechanische Verstärkung einen leistungsärmeren (und kostengünstigeren) Motor zu verwenden, um der aufgebrachten Kraft der Tellerfeder F_F entgegenzuwirken. Die Bolzenaufnahme (4) dient dem Zweck einer Kraftableitung in die Basis.
Sowohl Servomotor (BMH1003P36A1A) und Servoregler (Lexium32M32MD30N4) werden vom Hersteller Schneider Electric bezogen. Der Servomotor, besitzt eine Nenndrehzahl von 4000 U/min, ein Dauermoment von 8,4 Nm und eine mechanische Nennleistung von 2700 W am Ausgang [Schneider 2022]. Die Steuerung der Motoren (LMC058LF424) wird ebenfalls von Schneider Electric entwickelt. Sie besitzt eine Statusanzeige, zwei PCI-Steckplätze, USB Anschlüsse, E/A Module, einen RJ45-Anschluss für die serielle Schnittstelle RS232/RJ485, einen Schnittstelle für den Master Encoder und zwei CAN Schnittstellen. Im Zuge dieser Masterarbeit werden folgende Module verwendet: DM72F0 (7 Digitale Inputs, 2 Digitale Outputs), AI4LE (4 Analoge Inputs), DI12DE (12 Digitale In-puts).
In der Draufsicht, siehe Abb. 3, sind die angebrachten Sensoren schematisch eingezeichnet. Diese sind: (1) die Triangulationssensoren (Panasonic HG-C1050-P) zur Verschleißmessung, (2) die Kraftsensoren (HBM, Typ: KMR+ 400 kN) zur Belastungsmessung, (3) die Temperatursensoren (RS Pro Thermoelement, -75°C – 250 °C) zur Notabschaltung, (4) die induktiven Sensoren (XS4P18AB120) zur Zyklenzählung, für die Regelung der Kraft und Null-Position der Motoren, (5) die Impedanz Messung zur Detektion der galvanischen Trennung von Lager und Welle und (6) ein Schaltschrank, welcher die Messtechnik und Steuerungselektronik enthält.

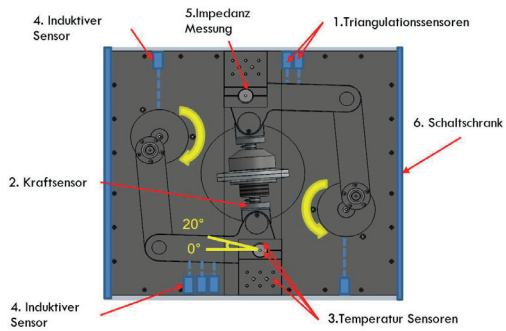


Abb. 3*:* Draufsicht Kniehebel Teststand

4. PRAKTISCHE DURCHFÜHRUNG

4.1 Erweiterung des Schaltschrankes und der Messtechnik

Im Zuge der Masterarbeit wird der Schaltkasten erweitert um die Bedienbarkeit zu erleichtern und um den Teststand zu automatisieren. Um Automatisierungsprozesse einzuführen, muss die bisher bestehende Kommunikationsstruktur adaptiert werden (siehe Abb. 4, rot gekennzeichnet), es wird eine Steuerung mit zusätzlichen Sensoren implementiert.

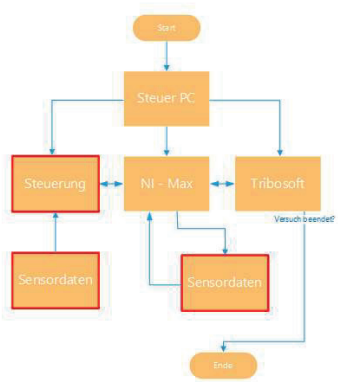


Abb. 4: Kommunikationsstruktur der Steuerung

Sowohl die Erfassung der Sensoren, als auch alle Voreinstellungen eines Versuches erfolgen über einen PCI – 6221 Connectorblock in der Software NI – Max. Diese Daten werden direkt an die AC²T firmeneigene Software „Tribosoft 1“, welche in LabVIEW programmiert und mit NI – Max kompatibel ist, weitergeleitet. Sie wird bei Teststänsden im Unternehmen eingesetzt um die Bedienbarkeit zu erleichtern und zu standardisieren. Die Tribosoft dient auch als Visualisierungstool während eines Versuches, in dieser sind alle Sensordaten die in NI – Max erfasst werden zu sehen. Nach Versuchsende exportiert Tribosoft alle aufgenommen Messdaten in Textfiles. Zu diesem bestehendem System werden die Steuerung und die in Rot gekennzeichneten Komponenten (siehe Abb. 5) implementiert.

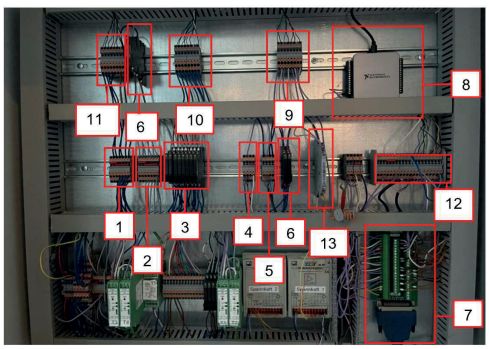


Abb. 5: Schaltschrank der Messelektronik
1. V+, 2. V-, 3. Schalter für die Steuerung, 4. Anschlüsse für die Impedanzmessung, 5. Anschlüsse Induktiven Sensoren 6. Schalter für Induktive Sensoren, 7. PCI – 6221 Connectorblock von NI, 8. USB – 6001 von NI ,9. Analoginput der Steuerung 10. Digital Inputs der Steuerung, 11. Digitalen Output der Steuerung 12. Anschlüsse für die Triangulationssensoren

Für den Ausbau des Schaltschrankes werden in erster Linie zwei Verteilerblöcke hinzugefügt, V+ (1) und V- (2). Diese bilden einen geschlossenen Stromkreis an denen alle neu hinzugefügten Komponenten, die eine Versorgung benötigen, angeschlossen werden. Dann werden die acht Halbleiterrelais (3) (Crydom DRSCN05) angeschlossen. Die Halbleiterrelais werden aus NI – Max mithilfe von Digitalen Ausgängen des PCI – 6221 Connectorblocks geschaltet (7). Der Ausgang der Halbleiterrelais führt zu den Digital Inputs der Steuerung (10), somit kann eine Reihe von Funktionen, die in der Steuerung programmiert sind, über NI – Max/Tribosoft genutzt werden. Weitere Halbleiterrelais werden ebenfalls in Kombination mit den im Teststand verbauten induktiven Sensoren verwendet (6). Erkennt ein induktiver Sensor ein Objekt, so schaltet das Relais. Das Ausgangssignal führt direkt in den Digitalen Output der Steuerung (11).
Das USB – 6001 von NI ist ein Multifunktions-Datenerfassungsgerät, welches zum bestehenden Tribometer hinzugefügt wird (8). Es besitzt Analog I/O, Digital I/O und einen Counter mit 32 Bit. Im Kniehebeltribometer werden für die Versuche nur der Analogoutputs (ein AO+ und ein AO-) verwendet. Diese führen direkt in den Analoginput (9) der Steuerung, sodass über den Steuer-PC eine Ausgangspannung definiert werden kann. Die Ausgangsspannung kann im weiterem Sinne einer proportionalen Einheit entsprechen, die in der Steuerung für die Automatisierung des Teststandes weiterverwertet werden kann. Das Gerät wird zusätzlich hinzugefügt, da keine übrigen AO auf dem PCI – 6221 Connectorblock zur Verfügung stehen.
Abb. 5 zeigt den ebenfalls neu hinzugefügten Optokoppler (13). Dieser wird in den Teststand integriert, da zum bestehendem Ausgang der Kraftsensoren in den NI – Connectorblock, noch ein Ausgang zur Steuerung geführt wird. Mit dem Optokoppler wird das Analogsignal in zwei galvanisch voneinander getrennte Systeme geschickt, da dies ansonsten zu einer galvanischen Kopplung führen kann und somit eines oder beide Messungen verfälscht. In weiterer Folge werden zusätzlich noch die Anschlüsse für die Impedanzmessung (4), die induktiven Sensoren (5) und die Triangulationssensoren (12) verbunden.

4.2 Steuerung eines Versuchlaufes

In dieser Arbeit wird die Software SoMachine von Schneider Electric verwendet. Antriebe von Schneider Electric können über die Steuerung mithilfe von SoMachine programmiert werden, mit dem Ziel eine Anwendung zu automatisieren. So wird auch in dieser Masterarbeit eine Reihe von Automatisierungsalgorithmen entwickelt, um einerseits die Bedienbarkeit des Teststandes zu erleichtern und andererseits die Versuchsgenauigkeit zu verbessern. In SoMachine wird eine graphische Struktur programmiert, in dem Blöcke seriell oder parallel geschaltet werden. Der Inhalt der Blöcke kann entweder eine graphische Programmierung mittels Funktionsbausteinen oder eine in Text geschriebene Ablaufsprache enthalten. Abb. 6 zeigt das Ablaufdiagramm eines spezifischen Versuches.

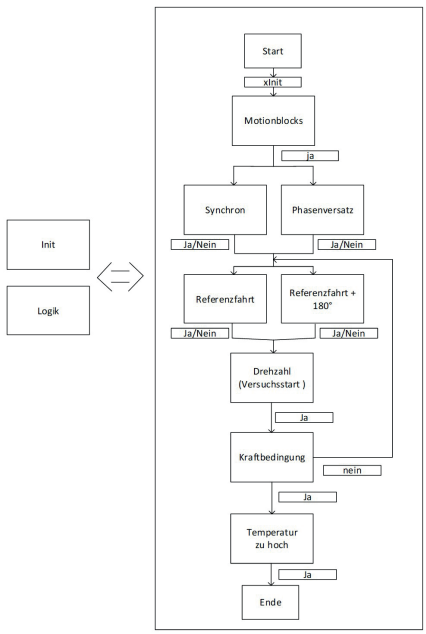


Abb. 6: Ablaufdiagramm eines Tribomenter Prüfstandes

Es wird von zwei Blöcken, die mit der Ablaufsprache programmiert werden ausgegangen (Init und Logik). Der erste Block „Init“ enthält unter anderem die Logik für die „xInit“-Bedingung und die Kommunikation zu den Servoreglern wird überprüft. Es wird hier ebenfalls die Drehzahlvorgabe und die Fahrtgeschwindigkeit zum Referenzpunkt festgelegt. Falls eine Kommunikation registriert wird, werden die Antriebe für einen Versuchsstart im „Motionblock“ freigegeben. Nun kann nach

betriebsbereiten Antrieben entschieden werden ob diese Positionssynchron oder mit einem Phasenversatz betrieben werden sollen.
Im nächsten Schritt können die Blöcke „Referenzfahrt“ und „Referenzfahrt + 130°“ ausgeführt werden. Ein zuvor sehr umständlicher Vorgang, bei dem auf den Antrieben physisch die Position geändert werden musste, ist nun automatisiert. Durch eine Kombination von Boolschen Operatoren und IF-Bedingungen sind die Antriebe programmiert in den unbelasteten und belasteten Zustand zu fahren. Dabei wird mit den Ausgängen der induktiven Sensoren (True/False), die Position der Antriebe gesetzt. Mit einem Taster wird die Tellerfeder entlastet (F=10 kN) und kann manuell gespannt werden. Der zweite Knopfdruck auf denselben Taster führt zu einer Rotation der Kurbelscheibe um 130° (F=Fmax [kN]) und die gespannte Kraft kann überprüft werden.

Der „Kraftbedingung“-Block wird in der Steuerung als Haltebedingung für einen Versuch verwendet, es wird in NI – Max eine Kraft vorgegeben. Bei Langzeitversuchen kommt es zu einer Lockerung der Tellerfeder. Der Grund dafür sind Vibrationen während des Versuches, welche zu einer Verdrehung der Tellerfeder führen können. Dies resultiert in einem Kraftabfall, ebenso entsteht ein Abfall der Kraft der Tellerfeder durch den Verschleiß der Lager. Wenn die Kraft um mehr als 5% sinkt, pausiert der Teststand. Dabei wird ein induktiver Sensor genutzt um die Kraft bei Maximaler Belastung zu überprüfen. Die Tellerfeder muss dann nachgespannt werden. Da die firmeneigene Software Tribosoft nicht dafür konzipiert ist, einen pausierten Versuch wieder zu starten, wird dies über einen Taster realisiert, der in der Steuerung programmiert wird um den Versuch vom letzten Haltepunkt wiederaufzunehmen.

4.3 Simulation

Onshape ist eine Cloud – basierte CAD Webapplikation, in dieser Arbeit wird sie zur Erstellung des Simulationsmodells benutzt [Onshape 2023]. Der Umweg über den Online-Dienst Onshape ermöglicht die Erstellung eines Simulationsmodelles mit geringerem Aufwand. Zunächst wird das in „SolidWorks“ erstellte CAD – Modell des Tribometers in Onshape importiert. Danach werden die Zwangsbedingungen in Onshape definiert. Mit MATLAB Simulink bzw. der integrierten Umgebung „Simscape Multibody“ wird anschließend die Mehrkörpersimulation durgeführt [MathWorks 2022]. Dabei wird ein Code in MATLAB geschrieben und Funktionen aus der Simscape Multibody Library aufgerufen. Das so generierte Simulink Modell ist zunächst nicht funktionsfähig und es müssen einige Änderungen (Zwangsbedingungen, Referenzpunkt, Koordinatentransformationen etc.) vorgenommen werden um ein für die Simulation adäquates System zu erstellen. Das globale Koordinatensystem wird neu verknüpft, überflüssig generierte Gelenke und Bezugsverknüpfungen werden gelöscht. Als Belastung wird eine Federkraft zwischen den Fixierungsaufbauten angebracht. Im nächsten Schritt werden Rampenfunktionen an den Wellen, die mit der Kurbelscheibe verbunden sind, definiert, um eine konstante Winkelgeschwindigkeit der Motoren zu simulieren. Anschließend werden zur Visualisierung Subsysteme mit Masken erstellt. Wird das Modell ausgeführt, ist das Mehrkörper – Model im Mechanics Explorer ersichtlich und führt eine zum Realsystem idente Bewegung aus. Mit Hilfe virtueller Messgeräte an den Gelenken und an der Feder ist es möglich mechanische Größen wie beispielsweise Kräfte oder Rotationen zu messen.

5. ERGEBNISSE

Im Laufe dieser Masterarbeit werden Gleitlager der Firma GGB vom Typ DX10 getestet. Diese werden für den Einsatz in Verschlusssystemen für Spritzgussformen entwickelt und es soll deren Lebensdauer unter Maximallast in Langzeitversuchen bestimmt werden. Insgesamt werden für diese Arbeit zwei Versuche ausgewertet, also 4 Lager getestet.
Der Verschleiß der Lager zu Versuchsende wird mittels Messung der Innengeometrie erfasst. Die im Zuge der Masterarbeit hinzugefügten optischen Sensoren werden mit diesen Verschleißdaten verglichen und analysiert – am Ende des Versuches sollten vergleichbare maximale Verschleißwerte wie bei der post-Test Messung erzielt werden. In den Ergebnissen wird auf zwei Versuche eingegangen und diese werden näher analysiert. Die von Tribosoft erzeugten Textfiles werden mit MATLAB weiterverarbeitet.

5.1 Langzeitversuch Lagerpaar 1

Der erste Versuch wird alle 100.000 Zyklen nachgeschmiert, läuft etwa 2,6 Millionen Zyklen (ca. 30 Tage bei 60/min), einer Maximallast von 150 kN bis die Temperatur in einem der Testlager auf über 50°C anstieg. Dann wird die Notabschaltung des Versuches ausgelöst. Nach Versuchsende werden die Lager ausgebaut und vermessen. In Abb. 7 und Abb. 8 sind die Differenzhöhen durch Verschleiß abgebildet. Lager 2 ist deutlich mehr verschlissen (mögliche Ursachen werden am Ende des Kapitel 5.2 diskutiert). Hier ist sowohl die Polymerschicht über die ganze Länge verschlissen, als auch die poröse Sinterschicht im unteren Teil des Lagers. Dies führt zu einem gemessenen maximalen Verschleiß von 0,75 mm. Das Lager 1 zeigt noch Reste der Polymerschicht über die gesamte Lagerlänge und der gemessene Verschleiß zeigt einen maximalen Betrag von 0,27 mm. Die Messdaten der Triangulationssensoren im Versuch 1 lieferten keine brauchbaren Ergebnisse. Der Sensor ist zu langsam um die Messdaten kontinuierlich während des Versuches zu erfassen. Die Bestimmung des Verschleißes über die Triangulationssensoren musste also im Versuch 2 angepasst werden.

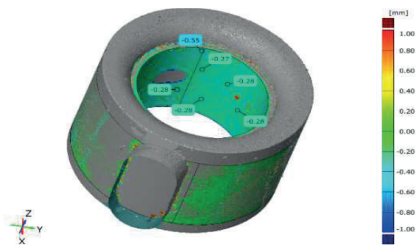


Abb. 7: Versuch 1, Höhendifferenz nach 3D-Vermessung von Lager 1 in [mm]

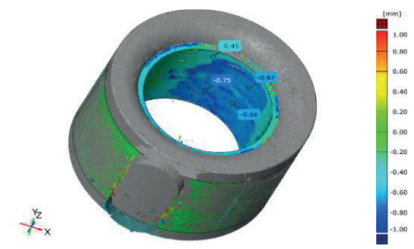


Abb. 8: Versuch 1, Höhendifferenz nach 3D-Vermessung von Lager 2 in [mm]

5.2 Langzeitversuch Lagerpaar 2

Der zweite Versuch wird erneut mit 60 U/min und mit einer Last 150 kN durchgeführt. Diesmal werden zum bestehenden Aufbau zwei zusätzliche Triangulationssensoren und zwei Impedanz Sensoren angebracht. Durch die Änderung des Stoppkriteriums bei einem Temperaturanstieg von 40 °C, wird dieser Versuch etwas früher beendet. (bei 1,85 Millionen Zyklen ca. 21 Tage). Im Versuch 2 ist auch die Impedanzmessung aktiviert. Somit ist es möglich den genauen Zeitpunkt einer ersten leitenden Verbindung zwischen Lager und Welle zu detektieren. In Abb. 9 ist ein real gemessener Teil des Kraftverlaufes von Versuch 2 abgebildet, es ist zu sehen das der Kraftverlauf des Realsystems sehr gut mit dem Simulationsmodell übereinstimmt.

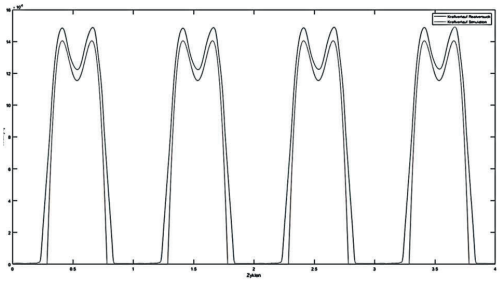


Abb. 9: Vergleich des Kraftverlaufes im Realsystem und im Simulationsmodell, Realversuch: Versuch 2, Lager 2: 1 Hz, 150 kN

In Abb. 10 und Abb. 11 sind die Verschleißbilder nach Ende des zweiten Versuches zu sehen. Wieder ist zu erkennen, dass bei einem Lager, im Vergleich zum anderen, der Verschleiß deutlich höher ist. Der Maximalwert liegt hier bei 0,9 mm Abtrag. Damit ist in beiden Versuchen das Lager 2 höher verschlissen.

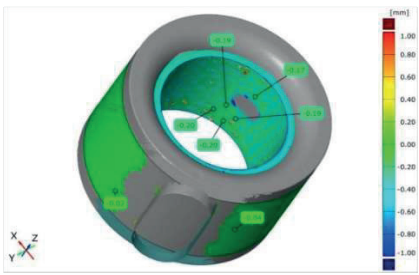


Abb. 10: Versuch 2, Höhendifferenz nach 3D-Vermessung von Lager 1 in [mm]

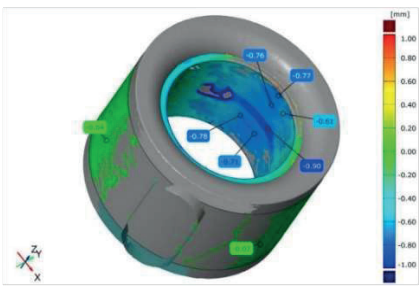


Abb. 11: Versuch 2, Höhendifferenz nach 3D-Vermessung von Lager 2 in [mm]

Da während des ersten Versuches die Abstandssensoren keine brauchbaren Ergebnisse lieferten, wird der Sensoraufbau adaptiert. Zu Beginn des Versuchs und auch in bestimmten Intervallen führt der Kniehebeltribometer nun eine Zustandskontrolle durch (circa alle 1.500 Zyklen, jede halbe Stunde), bei dem alle Messwerte der Sensoren überprüft werden, um eine Überschreitung von gewissen Maximalwerten auszuschließen. Für die Zustandskontrolle wird ein Zyklus mit 0,1 Hz gefahren. Mit der Reduktion der Versuchsgeschwindigkeit ist es nun möglich genauere Verschleiß-Messergebnisse zu erzielen. Die Triangulationssensoren werden bei diesem Versuch näher an der Bolzenaufnahme angebracht und es werden zwei Sensoren pro Seite eingesetzt, welche kontinuierlich den Abstand zum Hebel messen. Sensoren, die sich weiter entfernt von der Bolzenaufnahme befinden, liefern keine brauchbaren Messergebnisse. Die Messung auf den Hebel in der Schräglage führte zu ungenauen Messergebnissen.
In Abb. 12 und Abb. 13 ist der Verschleiß anhand der Triangulationssensoren zu sehen. Es ist anzumerken, dass nach dem Versuch nicht alle Daten verwertbar waren und deshalb einige Messwerte nicht abgebildet sind.

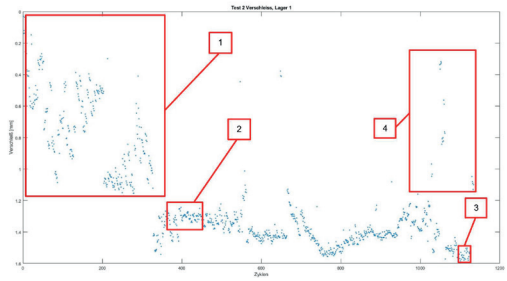


***Abb. 12**: Verschleißmessung, Versuch 2, Lager 1:1 Hz, 160 kN*

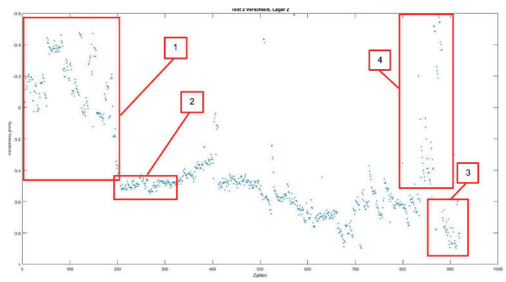


***Abb. 13**: Verschleißmessung, Versuch 2, Lager 2: 1 Hz, 160 kN*

Bis Konformität der beteiligten Reibpartner erreicht wird (Einlaufvorgang), ist eine höhere Streuung der Messwerte zu Beginn ersichtlich Bereich (1) in Abb. 12 und Abb. 13. Diese stabilisieren sich jedoch bei etwa 600.000 Zyklen in Bereich (2), der durch eine geringere Streuung charakterisiert ist. Jeweils zehn gemittelte Messungen nach dem Einlaufvorgang in Bereich (2) und gegen Ende des Versuches in Bereich (3) werden genutzt um den mittleren Verschleiß $\overline{r}$ zu berechnen:

$$\overline{r}_{Anf,Ende} = \frac{1}{n}\sum_{i=1}^{n} r_i \tag{5}$$

wobei r_i, die Messungen und n die Anzahl der Messungen sind. Anschließend wird die Differenz der mittelten Messwerte gebildet um auf den Verschleiß rückzuschließen:

$$\overline{r}_{Tot} = \overline{r}_{Ende} - \overline{r}_{Anf} \quad (6)$$

Die Verschleißmessungen ergeben eine Abnutzung von $\overline{r}$ =0.16 mm (ca. 0.2 mm mit Einlaufvorgang bei einer Annahme mit linearem Verschleißverlauf) bei Lager 1 und $\overline{r} = 0.45$ mm (ca. 0.6 mm mit Einlaufvorgang bei einer Annahme mit linearem Verschleißverlauf) bei Lager 2. Die Verschleißmessung von Lager 1 liegt am unteren Ende der nach dem Test ermittelten Werte. Der durch die Triangulationsmessdaten ermittelte Verschleiß von Lager 2 stimmt annähernd mit den post-Test 3D-Vermessungen überein. Durch den ungleichmäßigen Verschleiß kann der Maximalwert in der Mitte des Lagers leider nicht aufgelöst werden, da vom Triangulationssensor nur der Minimalwert der Lagerverschiebung gemessen werden kann. Am Ende des Tests zeigen sich die in Bereich (4) dargestellten Verschleißmessungen, die durch Ausreißer gekennzeichnet sind, eine „Zunahme“ des Abstandes zwischen Sensor und Hebel, also negativer Verschleiß. Speziell in treten diese Punkte nach den Nachschmierintervallen in einem bereits relativ hoch verschlissenen Lager auf. Es wird dazu folgende Hypothese aufgestellt: Verschleißpartikel können das geschlossene System nicht verlassen und setzten sich daher in den Schmiertaschen des Lagers ab. Durch das Nachschmieren werden diese zum Teil wieder frei gespült und können in die Kontaktzone gelangen. Dort erhöhen sie kurzzeitig wieder den Abstand zwischen Lager und Welle, bis sie wieder aus dem Kontakt gepresst werden. Dadurch kann es zu diesen kurzzeitigen negativen Verschleißwerten kommen.

Theoretisch wird ein symmetrisch, gleichförmiger Verschleißabtrag bei den Lagern erwartet. In beiden Versuchen ist die Lagerabnutzung jedoch bei einem Lager (Lager 2) größer im Vergleich zum anderen. Es ist anzunehmen das trotz einer nahezu symmetrischen Konstruktion, die Krafteinwirkung auf die Lager durch einen „asymmetrischen“ Zusammenbau nicht exakt symmetrisch ist. Dabei ist jeweils das Lager 2 von höherem Verschleiß betroffen. Hier ist nicht die betragsmäßige Krafteinwirkung der Lager gemeint, sondern die Krafteinwirkung der Lager in Längsrichtung. Wird ein Lager nicht flächendeckend gleichmäßig, sondern an einer Seite mehr beansprucht, so ist der Verschleiß folglich größer an dieser Stelle. Dadurch entsteht „punktuell“ eine frühzeitige Abnutzung der Polymerschicht, sodass es zu einem Metall-Metall Kontakt zwischen Lager und Welle kommt. Dies führt zu einer lokalen Temperaturerhöhung, die wiederum in einer Erweichung der Polymerschicht und einem höheren Verschleiß resultiert.

6. ZUSAMMENFASSUNG UND AUSBLICK

Im Zuge dieser Arbeit wurde das Verschleißverhalten von Gleitlagern im Kniehebel Tribometer erforscht. Dies bietet eine Möglichkeit die Abrasion von Lagern, zeitlich, unter diversen Bedingungen einzugrenzen und so hohe wirtschaftlich Kosten einzusparen.

Die im Rahmen dieser Masterarbeit durgeführten Tätigkeiten zur Analyse und Automatisierung des Teststandes führen Großteils zum Erfolg. Durch die Implementierung der Steuerung und der Sensorik werden Automatisierungsprozesse und neue Funktionen geschaffen, welche die Bedienung des Prüfstandes deutlich erleichtern. Ebenso wird die Versuchsgenauigkeit des gesamten Prüfstandes erhöht.

Das Simulationsmodell ist für die Analyse und das Verständnis des Kraftverlaufes des Kniehebeltribometers wichtig, es simuliert den genauen Kraftverlauf und an welchen Positionen mit welchen Kräften gerechnet werden muss.

Die implementierte Messeinrichtung um den Verschleiß online zu ermitteln, führte nur zu mäßig genauen Messergebnissen. Bei Lagern die hohen Verschleiß zeigten, werden durch die abgetragenen Verschleißpartikel im Kontakt Fehler in den Messergebnissen erzeugt. Die Messdaten der Lager mit geringem Verschleiß werden ausreichend genau erfasst. Die implementierten Sensoren erlauben nur eine Messung des minimalen Abstandes. Sollte ein Lager ungleichmäßig Verschleißen, kann das von einem Sensoren außerhalb des Lagers nicht erfasst werden.

Eine genauere Analyse dieser Messdaten und wie mit Messfehlern umgegangen werden kann ist Teil zukünftiger Untersuchungen. Zur Verbesserung der Verschleißanalyse mit den Triangulationssensoren wären in erster Hinsicht Sensoren mit höherer Messgenauigkeit erforderlich, als die derzeit verbauten. Derzeit wird der Versuchsstand Fett-geschmiert, Verschleißpartikel können daher nur unzureichend entfernt werden. Umlaufgeschmierte Systeme können hier Abhilfe schaffen, die die Verschleißpartikel im Ölkreislauf filtern. In weiterer Folge würden detailliertere Simulationen der Lager, um z.B. die thermische Ausdehnungen durch reibungsinduzierte Wärme analysieren, vorteilhaft sein. Zuletzt, um den Versuch weiter zu automatisieren, könnte eine Einrichtung, welche die Tellerfeder eigenständig nachspannt, geschaffen und in die Steuerung implementiert werden.

7. LITERATURVERZEICHNIS

Baumann, Peter. *Ausgewählte Sensorschaltungen.* Springer Fachmedien Wiesbaden, 2022.

Bernstein, Herbert. *Angewandte Leistungselektronik.* Springer Fachmedien Wiesbaden, 2021.

Electric, Schneider. „Benutzerhandbuch_LXM32M.“ 9 2017. (Zugriff am 5. 8 2022).

Horst, Czichos, und Karl-Heinz Habig. *Tribologie-Handbuch.* 65189 Wiesbaden, Deutschland: Springer, 2020.

Loose, Tobias. *Angewandte Regelungs- und Automatisierungstechnik.* Springer-Verlag, 2022.

MathWorks. *Simscape™ Multibody™ User's Guide.* 9 2022.

Onshape. 11. 1 2023. https://cad.onshape.com/help/Content (Zugriff am 15. 9 2022).

Probst, Uwe. *Servoantriebe in der Automatisierungstechnik.* Springer Fachmedien Wiesbaden, 2022.



Bony Vattappillil, Msc

AC2T-Research, Viktor-Kaplan-Straße 2/C
+43 2622/81600
bony.vattappillil@hotmail.de

PROZESSÜBERWACHUNG IM 3D BETONDRUCK: MESSUNG, AUFZEICHNUNG UND VERARBEITUNG VON PROZESSDATEN.

Völkl Janina

Kurzfassung: *Für ein bestehendes 3D Betondrucksystem wurde die Prozessüberwachung mittels Sensorik, Datenaufzeichnung und Datenverarbeitung verbessert. Für die Messung der Temperatur des Trockenmörtels, des Frischmörtels in der Düse und des Drucks in der Düse wurde geeignete Sensorik eingebracht. Die untersuchten Sensoren zur Messung des Wassergehalts im Frischmörtel erwiesen sich jedoch als nicht geeignet für das bestehende System. Zur Verbesserung der Qualitätssicherung wurden die Messdaten aufgezeichnet und in ein übersichtliches Druckprotokoll verarbeitet, welches sich über eine Webseite downloaden lässt. Über diese Webseite lassen sich die Messdaten zudem visualisieren. Der mittels statistischer Versuchsplanung (DoE) untersuchte Bereich der Frischmörtelfeuchte und des Beschleunigergehalts erwies sich als passend für das untersuchte 3D-Betondrucksystem. Für mehr Reihen bis zum Einbruch eines definierten Druckobjekts erwies sich ein geringer Materialfeuchte- und hoher Beschleunigeranteil als vorteilhaft. Auf die Biegezug- und Druckfestigkeit hat ein geringer Beschleunigeranteil signifikante Auswirkungen gezeigt.*

Schlüsselwörter: *3D Betondruck, Sensorik, Qualitätssicherung, statistische Versuchsplanung.*

1. EINLEITUNG

Der 3D Druck ist generell eine stark wachsende Branche. Dies zeigt sich auch in der Bauindustrie, in der immer mehr im Bereich des 3D Betondrucks entwickelt und geforscht wird. Das Spektrum des 3D Betondrucks ist hierbei groß und reicht vom Drucken von Kunstobjekten bis hin zum Druck von ganzen Häusern [Vasilic et al. 2022].

Was die verschiedenen Einsatzgebiete des 3D Betondrucks gemeinsam haben, sind die hohen Materialansprüche an den Druckbeton und an das Betondruckobjekt. Um diese Materialansprüche zu erfüllen, sind jedoch nicht nur die verwendeten Materialien relevant, sondern auch die korrekte Verarbeitung und vor allem das Mischungsverhältnis zwischen Trockenbeton, Wasser und Beschleuniger. Beschleuniger kommen im 3D Betondruck häufig zum Einsatz, um die Aushärtezeit des Betons zu verkürzen. Aufgrund dessen ist es sinnvoll, relevante Prozessparameter (z.B.: Temperaturen und Drücke) zu messen, aufzuzeichnen und die Daten passend für die Qualitätssicherung und Prozessüberwachung zu verarbeiten. Ist der Frischbeton nämlich nicht passend gemischt, kann er zu stark verrinnen oder er ist zu fest, um sich gut mit den anderen Lagen zu verbinden. Um diese Anforderungen gut zu erfüllen, benötigt es Personal mit viel Erfahrung um die passenden Einstellungen vorzunehmen und das System während des Druckprozesses zu überwachen. Durch Verbesserungen der Prozessüberwachung kann es hingegen auch weniger erfahrenem Personal leichter gemacht werden den Druckprozess passend einzustellen und zu überwachen. Zudem kann eine gute Prozessüberwachung als Basis für eine darauf aufbauende Regelung einzelner Prozessparameter genutzt werden.

2. PROBLEM- UND AUFGABENSTELLUNG

Ziel dieser Masterarbeit ist es, ein bestehendes 3D Betondrucksystem, zur Adressierung der oben genannten Probleme, mit entsprechender Sensorik auszustatten. Dazu müssen zuerst relevante Systemparameter identifiziert werden. Anschließend sollen passende Sensoren ausgewählt und in das System eingebunden werden. Dabei gilt der Findung und Prüfung der Sensorik zur Messung der Materialfeuchte des Frischmörtels besonderes Augenmerk. Im Sinne der Qualitätssicherung ist es erforderlich, sowohl die Messdaten der neu eingebundenen Sensorik, als auch der bereits vorhandenen Sensorik und relevanter Systemparameter aufzuzeichnen. Diese Daten werden in einem Druckprotokoll verarbeitet und auf einer benutzerdefinierten Webseite der SPS (Speicher programmierbare Steuerung) visuell dargestellt und zum Download bereitgestellt. Zudem sollen die Auswirkungen der Materialfeuchte- und des Beschleunigergehalts auf die Anzahl der druckbaren Reihen bis zum Einbruch des Objekts, sowie die Biegezug- und Druckfestigkeit untersucht werden. Die Zusammenhänge sollen mittels einer geeigneten Formel beschrieben und die Qualität des Modells anhand statistischer Größen und den für die statistische Versuchsplanung typischen Diagrammen bewertet werden.

3. PRAKTISCHE DURCHFÜHRUNG

Der praktische Teil gliederte sich in die Implementierung der Sensorik, die Prozessdatenverarbeitung und Messdatenvisualisierung und die statistische Versuchsplanung.

3.1 Sensorik

Für die Messung des Frischmörtelfeuchtegehalts wurden ein Time Domain Reflectometry Sensor, kurz TDR-Sensor und ein Mikrowellen-Resonanz-Sensor, welche beide auf einem dielektrischen Messprinzip beruhen, untersucht.

Zur Messung der Frischmörtelfeuchte erfordern beide Sensoren eine materialspezifische Kalibrierkurve. Zur Erstellung dieser Kalibrierkurve wurden für beide Sensoren Messungen bei einem massebezogenen Feuchtegehalt von 13,5%; 14,5% und 15,5% durchgeführt. Diese Messreihe wurde randomisiert drei Mal mit bei Raumtemperatur (~22°C) gelagertem Trockenmörtel und zwei Mal bei mit bei 5°C gelagertem Trockenmörtel durchgeführt. Für beide Sensoren konnten jedoch keine zufriedenstellenden Kalibrierkurven gefunden werden. Mit dem untersuchten Mikrowellen-Resonanz-Sensor zur Materialfeuchtemessung konnte ein Standardfehler von rund 0,35 Prozentpunkten Materialfeuchte erreicht werden, was für den Einsatz bei einem Feuchtebereich des Frischmörtels von etwa 13,5 bis 15,5% Materialfeuchte nicht ausreichend ist. Der Standardfehler des TDR-Sensors war mit 0,59 sogar noch höher. Die große Abweichung (hellblau) zwischen den tatsächlichen Feuchtewerten (orange) und den Schätzwerten des Modells (rot) lässt sich in Abb. 1 erkennen.

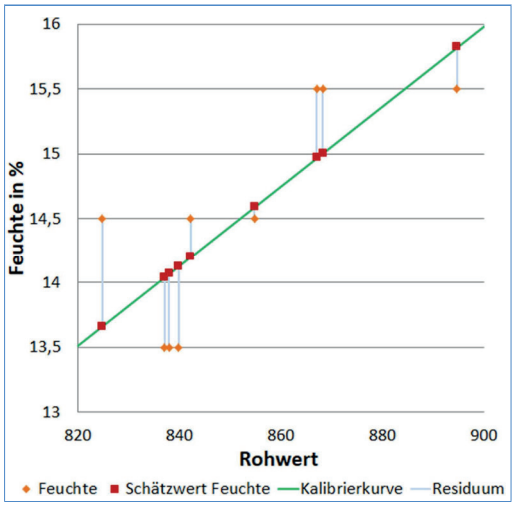


Abb. 1: Kalibrierkurve, massebezogene Frischmörtel Materialfeuchte und massebezogene Frischmörtel Materialfeuchte Schätzwerte des TDR-Sensors mit bei Raumtemperatur (~22°C) gelagertem Trockenmörtel T1

Eine Berücksichtigung der Frischmörteltemperatur führte zu einer Verschlechterung des Modells.

Für den Einbau des Sensors in den Mörteltrog erwies sich der TDR-Sensor zudem als gänzlich ungeeignet, da es hier zu extremen Schwankungen der Messwerte kam (siehe Abb. 2).

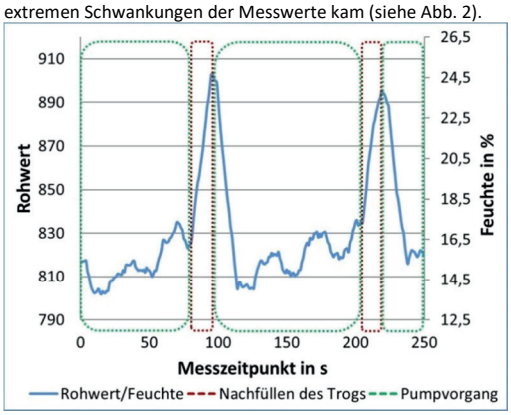


Abb. 2: Verlauf des einheitenlosen Rohwertes des TDR-Sensors und der gemessenen massebezogenen Frischmörtelmaterialfeuchte im Trog

Somit konnten lediglich für das System geeignete Sensoren zur Messung der Trockenmörteltemperatur, der Frischmörteltemperatur in der Düse und des Drucks in der Düse in das System eingebracht werden (siehe Abb. 3)

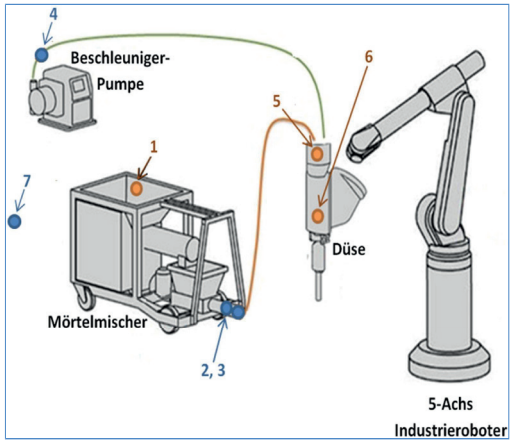


Abb. 3: Einbaustellen der neu eingebauten Sensorik (orange) und der bereits zuvor vorhandenen Sensorik (blau), 1: Trockenmörteltemperatur, 2: Frischmörteltemperatur am Schlauchanfang, 3: Druck am Schlauchanfang, 4: Beschleunigerdruck, 5: Druck vor der Düse, 6: Temperatur in der Düse, 7: Umgebungstemperatur

3.2 Prozessdatenverarbeitung

Zuerst wurde auf der SPS eine Aufzeichnung aller Messdaten, sowie einiger anderer Druckobjekt spezifischer Einstellungen, wie zum Beispiel der Förderrate der Mörtelpumpe und Chargennummern, implementiert.

Anschließend wurde auf der SPS eine benutzerdefinierte Webseite programmiert, welche die aufgezeichneten Daten in ein Druckprotokoll verarbeitet, zum Download bereit stellt und es ermöglicht, die Messdaten als yt-Diagramm zu visualisieren. Das Druckprotokoll ist für die Qualitätssicherung besonders wichtig und gibt eine Übersicht über alle gedruckten Objekte. Es listet auf der Übersichtsseite alle Druckobjekte inklusive den Eingaben des Users, wie unter anderem dem Beschleunigeranteil und der verwendeten Chargennummer, so wie die Mittelwerte der Messgrößen auf. Auf den folgenden Seiten des Druckprotokolls lässt sich der genaue Verlauf aller Messgrößen des 3D Betondrucksystems einsehen.

Durch die Möglichkeit der Visualisierung aller Messdaten über die von der SPS bereitgestellte Webseite kann insbesondere bei fehlerhaften Drückobjekten leicht geprüft werden, ob es während dem Druck des Objekts ungünstige Messgrößenverläufe gab.

3.3 Statistische Versuchsplanung/Design of Experiment (DoE)

Mittels statistischer Versuchsplanung wurden die Auswirkungen des Beschleunigeranteils und der Frischmörtelfeuchte auf die druckbaren Reihen bis zum Einbruch eines definierten Druckobjekts und die Biegezug- und Druckfestigkeit nach 28 Tagen untersucht. Hierfür wurde ein voll faktorieller 2-Level Versuchsplan inklusive Mittelpunkt gewählt. Der für das System empfohlene Beschleunigerbereich von 4-6% und Frischmörtelfeuchtegehalt von 14-15,5% wurden jeweils als obere und untere Grenze verwendet. Dabei wurde der Frischmörtel manuell, mit dem entsprechenden Feuchtegehalt, angemischt und der Beschleunigeranteil wurde am System eingestellt. Es wurden drei Versuchsreihen mit randomisierter Testreihenfolge durchgeführt um die Auswirkungen von zufälligen Schwankungen auf die Ergebnisse zu minimieren.

Die Auswertung erfolgte mittels Design-Expert. Der mittels des Programms gefundene Zusammenhang ergab sich zu

$$R = 44{,}36 - 3{,}11 * h_{\mathrm{m}} + 3{,}67 * B \quad (1)$$

$$f_{\mathrm{ct,28d}} = 12{,}24 - 0{,}42 * \mathrm{B} \quad (2)$$

$$f_{\mathrm{ck,28d}} = 64{,}62 - 4{,}07 * \mathrm{B} \quad (3)$$

wobei R den druckbaren Reihen bis zum Einbruch, $f_{\mathrm{ct,28d}}$ der Biegezugfestigkeit nach 28 Tagen in MPa, $f_{\mathrm{ck,28d}}$ der Druckfestigkeit nach 28 Tagen in MPa, h_{m} der massebezogenen Materialfeuchte in % und B dem volumenbezogenen Beschleunigeranteil in % entspricht.
Es hat sich gezeigt, dass die Frischmörteltemperatur keine signifikanten Auswirkungen hat. Auf die druckbaren Reihen bis zum Einbruch hatten dadurch lediglich der Beschleunigeranteil und die Materialfeuchte signifikante Auswirkungen. Die Abhängigkeit der Reihen bis zum Einbruch von der Materialfeuchte und dem Beschleunigeranteil lässt sich in Abb. 4 erkennen.

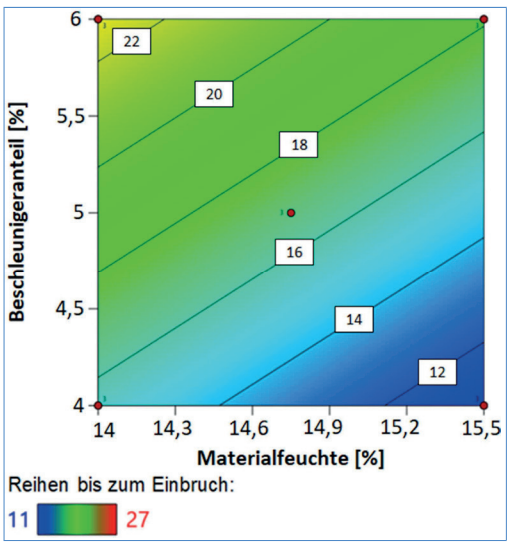


Abb. 4: Abhängigkeit der Reihen bis zum Einbruch von der Materialfeuchte und dem Beschleunigeranteil (Bedeutung der Farben: siehe Legende unter der Graphik)

Auf die Biegezug- und Druckfestigkeit nach 28 Tagen hatte sogar nur der Beschleunigeranteil signifikante Auswirkungen. Das Modell hat gezeigt, dass ein höherer Beschleunigeranteil und ein niedrigerer Frischmörtelanteil förderlich sind, um mehr Reihen zu Drucken, bevor es zu einem Einbruch des Druckobjekts kommt. Von einer Reduktion der Materialfeuchte auf unter 14% ist jedoch für das untersuchte System abzuraten, da es sonst beim Pumpen des Frischbetons häufiger zu Problemen kommen kann. Auch eine Erhöhung des Beschleunigeranteils über die Grenze von 6% ist nicht zu empfehlen, weil es bei solch einer Erhöhung oft zu Inhomogenität der Druckbahn kommt. Zudem ist zu bedenken, dass ein höherer Beschleunigeranteil stets zu einer geringeren Biegezug- und Druckfestigkeit führt.
Zur Beurteilung der Modelle wurden insbesondere die für DoE typischen Diagramme zur graphischen Auswertung herangezogen. So lassen sich zum Bespiel mittels Cook's Distanz Diagramm Ausreißer anhand ihrer Cook's Distanz leicht erkennen. Werte mit einem Wert höher als 1 gelten als kritisch und verfälschen die gesamte Vorhersage. In der Regel gilt eine Cook's Distanz von über 0,5 als potenziell problematisch.
Die Diagramme der Cook's Distanz aller Modelle weisen keine kritischen Ausreißer, und allgemein niedrige Cook's Distanzen auf. Lediglich bei den Reihen bis zum Einbruch ist der 13te Messwert etwas hoch, ein Wert von knapp 0,4 (siehe Abb. 5), ist jedoch nicht weiter bedenklich. In Bezug auf Ausreißer wird das Cook's Distanz Diagramm oft in Kombination mit dem DFFITS vs Run Plot betrachtet, weil jener Plot Aufschluss über die Auswirkungen einzelner Messpunkte auf die Vorhersage gibt. Auch hier lieferten alle Modelle zufriedenstellende Ergebnisse.
(vgl. [Siebertz et al. 2017])

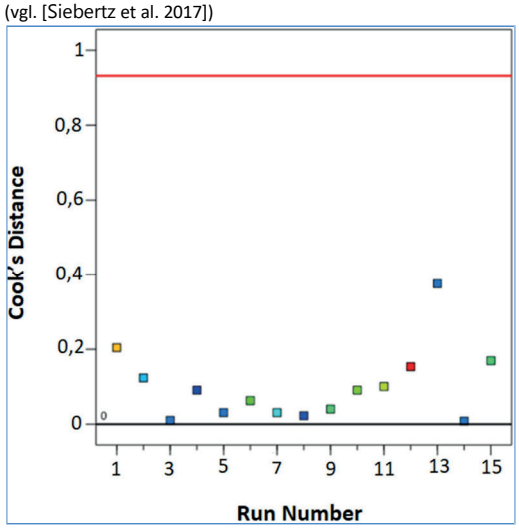


Abb. 5: Cook's Distanz Plot bei Reihen bis zum Einbruch

6. ZUSAMMENFASSUNG UND AUSBLICK

Die Prozessüberwachung und Qualitätssicherung eines bestehendes 3D-Betondrucksystem konnte erfolgreich durch die Integration geeigneter Sensorik, deren Datenaufzeichnung und Verarbeitung, so wie die Untersuchung der Auswirkungen des Beschleunigeranteils und Frischmörtelfeuchtegehalts verbessert werden.
Für die Messung des Feuchtegehalts des Frischmörtels konnte keine für das System geeignete Sensorik gefunden werden. Sowohl der Standardfehler des untersuchten TDR- als auch des Mikrowellen-Resonanz-Sensors erwies sich als nicht zufriedenstellend. Der TDR-Sensor erwies sich zudem als nicht geeignet für den Einbau in den vorhandenen Mörteltrog, weil die Messergebnisse hier extrem geschwankt haben.
Für die Qualitätssicherung besonders förderlich ist zudem die Implementierung eines sich automatisch generierenden Druckprotokolls. Dieses lässt sich über eine Webseite herunterladen und bietet eine Übersicht über alle gedruckten Objekte. Das Druckprotokoll beinhaltet wichtige Parameter wie unter anderem Chargennummern, Förderleistung und die aufgezeichneten Messwerte. Zudem lassen sich die aufgezeichneten Messwerte über die erstellte Webseite visualisieren.
Mittels DoE konnten geeignete Formeln für die Reihen bis zum Einbruch eines definierten Druckobjekts, sowie für die Biegezug- und Druckfestigkeit nach 28 Tagen gefunden werden. Hierbei hat sich gezeigt, dass bei den Reihen bis zum Einbruch sowohl der Beschleunigeranteil, als auch der Feuchtegehalt des Frischmörtels signifikante Auswirkungen

hatten. Auf die Biegezug- und Druckfestigkeit nach 28 Tagen hatte jedoch nur der Beschleunigeranteil signifikante Auswirkungen, der Materialfeuchtegehalt lieferte für den untersuchten Bereich jedoch keinen signifikanten Einfluss. Die Frischmörteltemperatur hatte bei allen Modellen keinen signifikanten Einfluss.

Die Integration von mehr Sensorik und die Aufzeichnung der Messdaten bildet die perfekte Basis, um Datenanalysen durchzuführen und Zusammenhänge zwischen einzelnen Messgrößen und deren Auswirkungen auf das Druckobjekt zu identifizieren. Aufbauend auf diesen Analysen und den im Rahmen dieser Masterarbeit untersuchten Auswirkungen des Beschleunigeranteils und der Frischmörtelfeuchte auf die Biegezug- und Druckfestigkeit sowie Reihen bis zum Einbruch des Druckobjekts, könnte zudem eine Regelung verschiedener Parameter implementiert werden. Dadurch wäre es unter anderem möglich, das 3D Betondrucksystem besser an verschiedene Umgebungsbedingungen, wie Umgebungstemperatur und Luftfeuchtigkeit anzupassen, um dadurch insbesondere beim Drucken im Freien potenziell bessere Ergebnisse zu erzielen. Bezüglich der Einbringung weiterer Sensorik ist es sinnvoll insbesondere Sensoren zur Messung des Frischmörtelfeuchtegehalts weiter zu untersuchen um eine geeignete Lösung zur Messung der Frischmörtelfeuchtigkeit zu finden. Auch eine in-line Messung der Viskosität des Frischmörtels wäre sehr vorteilhaft, weshalb der Markt für Viskositätssensoren im Auge behalten werden sollte um einen Viskositätssensor zu finden, welcher sich auch für die in-line Messung von Frischbeton eignet.

7. LITERATURVERZEICHNIS

Vasilic, K.; Hack, N.; Kloft, H.; Lowke, D.; Mechtcherine, V.; Nerella, V.; Wangler, T., 2022. Digitale Fertigung im Betonbau. BetonKalender 2022, 68, 692-724, 9783433033449

Siebertz, K.; van Bebber, D.; Hochkirchen, T., 2017. Statistische Versuchsplanung: Design of Experiments (DoE)., Springer Vieweg, 978-3-662-55742-6, Berlin, Heidelberg



Janina Völkl B.Sc.

janina_voelkl@live.at

2021-	Softwareentwicklerin bei Baubot
2019-2023	Masterstudiengang Mechatronik Fachhochschule Wiener Neustadt
2016-2019	Bachelorstudiengang Mechatronik|Mikrosystemtechnik Fachhochschule Wiener Neustadt



FACHHOCHSCHULE WIENER NEUSTADT

University of Applied Sciences – Austria

ISBN 978-3-200-09659-2



Fachhochschule Wiener Neustadt GmbH
University of Applied Sciences

Johannes Gutenberg-Straße 3, 2700 Wiener Neustadt, Austria
+43 (0) 26 22 | 89 0 84 - 0 | office@fhwn.ac.at | fhwn.ac.at

Herausgegeber: Studiengang Mechatronik
Foto-Credits © FH Wiener Neustadt GmbH, BMI

PROCEEDINGS 2023 | Mechatronik | Mechatronics