iCCC 2026

iCampus Cottbus Conference

Transferkonferenz für Innovation in Wissenschaft und Industrie

Tagungsband

Sensorik, MEMS & KI als Schlüsselfaktoren im Strukturwandel

https://**iCCC2026.de**

iCampus COTTBUS

Gefördert durch:

Bundesministerium für Forschung, Technologie und Raumfahrt

Tagungsband

iCampus Cottbus Conference 2026

Bei diesem Band handelt es sich um den Kongressband der
zweiten iCampus Cottbus Conference 2026.
Dieser Band beinhaltet die Manuskripte zu den jeweiligen Vorträgen und
Posterbeiträgen.

Für Form und Inhalt der Beiträge zeichnen sich die Autoren verantwortlich.
Die AMA Service GmbH übernimmt keine Gewähr für die Richtigkeit, Genauigkeit
und Vollständigkeit der Angaben sowie die Beachtung privater Rechte Dritter.

Herausgeber / Publisher:

AMA Service GmbH
Von-Münchhausen-Str. 49
31515 Wunstorf / Germany
Tel. +49 5033 9639-0
info@ama-service.com
www.ama-service.com / www.sensor-test.com
www.ama-science.org / www.smsi-conference.com

ISBN 978-3-910600-10-2

© AMA Service GmbH, 2026

Sehr geehrte Damen und Herren, liebe Projektpartner, verehrte Gäste,

im Zuge des Strukturwandels in der Lausitz wurden seit 2020 vielfältige Aktivitäten unternommen und erste bedeutende Beiträge zur Stärkung der Innovations- und Wettbewerbsfähigkeit der Region geleistet. Beispiele stellen die Eröffnung Europas modernsten ICE-Bahninstandhaltungswerks in Cottbus, die Gründung der Medizinischen Universität Lausitz oder die durch Förderung initiierte intensivierte Zusammenarbeit von Wissenschaftseinrichtungen und vor allem regionalen Unternehmen dar. Der Strukturwandel nimmt Fahrt auf, und wir sind nicht nur mittendrin, sondern haben auch maßgebliche Mitgestaltungsmöglichkeiten.

Die iCampµs Cottbus Conference soll in diesem Kontext Impulse zur Ideenfindung und Vernetzung setzen und die Potenziale leistungsfähiger Sensorik, innovativer Mikroelektronik und Künstlicher Intelligenz verdeutlichen – sei es für Industrie 4.0, eine moderne Energiewirtschaft oder die (Tele-)Medizin. Das große Interesse an der erstmals im Jahr 2024 durchgeführten biennalen Konferenzreihe unterstreicht die einzigartige Bedeutung für die Lausitzer Region und darüber hinaus: Mit 230 Teilnehmenden aus Wirtschaft, Forschung und Verwaltung, blicken wir auf einen erfolgreichen Auftakt mit überregionaler Strahlkraft zurück.

An diesen Erfolg möchten wir im Jahr 2026 anknüpfen und die Konferenz weiter ausbauen. Im Vergleich zur iCCC2024 konnten wir eine Zunahme eingereichter Full Paper um 50% verzeichnen. Neben Fachbeiträgen, Posterpräsentationen und Begleitausstellung werden in diesem Jahr erstmalig auch zwei Workshops angeboten. Besonders hervorheben möchte ich den Gründungsworkshop „Entrepreneurship“. Gründerinnen und Gründer spielen eine Schlüsselrolle für den Strukturwandel. Mit ihren innovativen Ideen und Geschäftsmodellen setzen sie neue Impulse und schaffen Arbeitsplätze. Damit sind Sie ein wichtiger Baustein, um die Wettbewerbsfähigkeit der Region nachhaltig zu stärken. Denn der Strukturwandel kann nur durch unser aller Einsatz gelingen. Packen wir diese Herausforderung gemeinsam an!

In diesem Sinne lade ich Sie herzlich zur iCampµs Cottbus Conference 2026 ein. Insgesamt erwartet Sie ein vielfältiges Programm und ich freue mich auf interessante Fachbeiträge und anregende Diskussionen mit Ihnen!

Prof. Dr. Harald Schenk
Gesamtprojektleiter iCampµs Cottbus und Chairman iCCC2026

iCCC 2026 iCampus Cottbus Conference

Programm

Di. 05.05.2026

16:00 – 18:00 Registrierung und Welcome Reception im Konferenzhotel Radisson Blu
17:00 – 19:15 Führungen durch Europas modernstes ICE-Bahninstandhaltungswerk

1. Tag / Mi. 06.05.2026

ab 08:00 Registrierung
09:00 – 09:15 Eröffnung: BTU-Präsidentin **Prof. Dr. Gesine Grande;** Grußworte: BMFTR **Dr. Oliver Höing,** MWFK **Staatssekretär Tobias Dünow**
09:15 – 10:15 **KEYNOTES: Rolf Härdi,** CTIO Strategie, Innovation und neue Technologien, Deutsche Bahn AG; **Dr. Gabriel Kittler,** CEO X-Fab Erfurt
10:15 – 10:45 Ausstellerpitches
10:45 – 11:15 Kaffeepause und Ausstellung

11:15 – 13:00 **SESSION I: Condition Monitoring**
Invited: **Dr. Konstantin Jonas,** Deutsche Bahn AG
I.1: Lehmann (Fraunhofer IVI): Field Data optimized Lifetime Prediction of Li-Ion Batteries in Electric Vehicle Fleets
I.2: Liebermann (Fraunhofer IIS): Smart Sensor for continuous condition monitoring and predictive maintenance of a milling machine
I.3: Kästle (BAM): Berstdruckvorhersage von faserverstärkten Druckbehältern anhand von Schallemissionen

SESSION II: Gesundheit und Sport
Invited: **Dr. Barlo Hillen,** Sportmedizin Uni Mainz
II.1: Schlierkamp-Voosen (Sportmedizin Uni Mainz): Integrative Bewertung von Mikrozirkulationsstörungen bei Post-Covid 19 mittels Hyperspektralbildgebung in Kombination mit kardiopulmonaler Leistungsdiagnostik
II.2: Steinmetzer (BTU): Mobility Monitoring for Geriatrics: Gait Detection and Parameter Extraction using an e-Textile Garment IMU and Deep Learning
II.3: Schäfer (Fraunhofer IMM): Implantable MEMS Sensor Platform for Measuring Heart Activity
II.4: Burmann (Fraunhofer ISST): Trainingssteuerung im Leistungssport: Integration multimodaler Daten durch Datenraumtechnologien

13:00 – 14:00 Mittagspause und Ausstellung

14:00 – 16:00 **SESSION III: Workshop Entrepreneurship**
Dr. Andreas Olmes,
Lena Strobl/Fiberflow,
Dr. Mark Kuprat/e-phant,
Team des FG **Prof. Mißler-Behr/**BTU
Robert Holzschuh/MUL-CT
Ablauf: · Einführung · 2 Praxisbeispiele · Podiumsdiskussion unter Einbeziehung des Publikums · Startup Lausitz Accelerator (ein Projekt der Thiem-Research GmbH/ Medizinische Universität Lausitz-Carl Thiem) · Gründungsservice BTU: „Start-Up Factory“

SESSION IV: Energiewirtschaft und Material & Prozesstechnologie
Invited: **Dr. Thomas Porsche,** Standortentwickler LEAG
IV.1: M. Schenk (BTU): Hochspannungstechnik & Elektrische Anlagen
IV.2: Frenzel (HS Zwickau): Untersuchung der optischen Eigenschaften von Palladiumdünnschichten in Wasserstoff für einen Chip-basierten MOEMS-Wasserstoff-Sensor
IV.3: Bock (BAM): Voll digitalisiertes Reallabor Wasserstofftankstelle der BAM
IV.4: Nobari: Portable NIR Spectroscopy for Reliable Characterization of High-Intensity Radiating Objects: Statistical vs AI-Based Modelling
IV.5: Ernst (BTU): Halbleiter für Quantumcomputing: Wertschöpfung und chemischer Fingerabdruck von isotopenreinem Silicium-28
IV.6: Rosencrantz (Fraunhofer IAP & BTU): Mehr als Bionik – Biofunktionelle Polymere als innovative neue Materialien

16:00 – 16:30 Kaffeepause und Ausstellung
16:30 – 17:30 Postersession und Rundgang der Jury
17:30 – 18:45 Sciences Slams „7 in 49“ im Pücklersaal: Warming up: Überraschungsgast
19:00 – 22:00 Conference Dinner

2. Tag / Do. 07.05.2026

ab 08:00 Registrierung
08:30 – 09:00 Begrüßung und Einführung **Prof. Dr. Harald Schenk**
09:00 – 10:00 **KEYNOTES: Prof. Dr. Michael Totzeck,** ZEISS; **Dr. Michael Hosemann,** Siemens Healthineers
10:00 – 10:30 Kaffeepause und Ausstellung

10:30 – 12:00 **SESSION V: Bio- und Umweltsensorik**
Invited Talks: **Dr. Attila Michael Bilgic,** CEO KROHNE Group;
Prof. Dr. Ulrike Wallrabe, University of Freiburg, IMTEK
V.1: Reiter/Wenger/Fischer (BTU, IHP):
Entwicklung von Plasmonischen On-Chip Sensoren
V.2: Ulbricht (FBH): Auf dem Weg zur radarbasierten Rinderüberwachung: Erste Überlegungen, Aufbauten und Ergebnisse
V.3: Richter (TU DD): Mehrstufenprotokolle zur Diagnose von Atemwegs-infektionen automatisiert mit chemofluidischen Schaltkreisen

SESSION VI: Kommunikation
Invited: **Dr. Karsten Walther,** CEO Perinet
VI.1: Schmieder (Fraunhofer IPMS): The GigaCoupler – a scalable Opto-Isolator with Gpbs-Data Transmission
VI.2: Farmlink: Fabia Galka, Milan Ferus-Comelo, Daniel Strauß, Duale Hochschule Baden-Württemberg (DHBW): Dezentrale, Mesh-basierte Plattform zur Überwachung von Nutztieren für ländliche Gebiete mit eingeschränkter Konnektivität
VI.3: Krause/Näser (BTU): RFID und BLE Tracking im industriellen Umfeld

12:00 – 13:30 Mittagspause und Ausstellung

13:30 – 15:00 **SESSION VII: Ultraschallsensorik und Mobilität**
Invited: **Dr. Eric Starke,** Endress&Hauser, Sick
VII.1: Jongmanns/Monsalve (Fraunhofer IPMS): 1D phased array for imaging applications in air utilizing MEMS ultrasonic transmitters
VII.2: Lützelberger/Drese (HS Coburg): Ultrasound-Based Interferometric Hip Implant Monitoring: Analysis of Frequency Selection and Minimum Detectable Gap Width
VII.3: Kramer (DLR): Modellierung einer Heat-Pipe-gestützten Thermal-regulierung von HT-PEM-Brennstoffzellen in der Luftfahrt – Numerische Bestimmung geometriespezifischer Wärmeübergangskoeffizienten

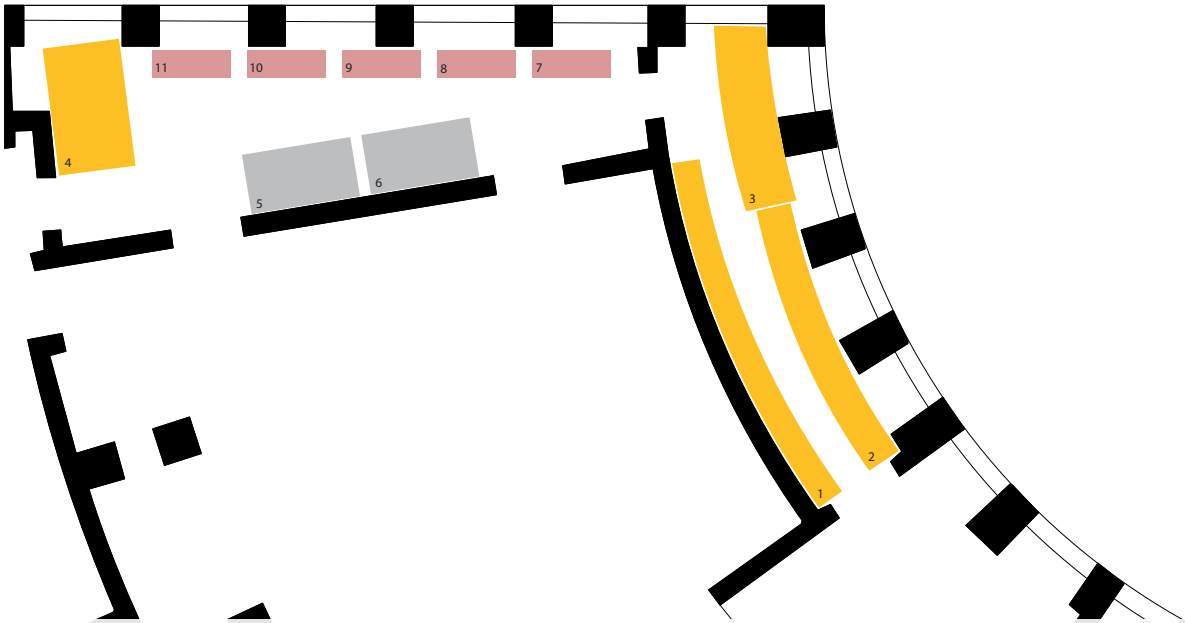
SESSION VIII: Workshop 5G
mit Boldyn & Nokia, DB, 5G-Team BTU
Live-Demonstrationen im 5G-Netz

15:00 – 15:30 Preisverleihung: Best Paper, Best Poster, Best Booth

iCampus COTTBUS

Gefördert durch:
Bundesministerium für Forschung, Technologie und Raumfahrt

iCCC 2026



ACTEMIUM

b-tu Brandenburgische Technische Universität Cottbus - Senftenberg

chesco Center for Hybrid Electric Systems Cottbus

DLR Institut für CO_2-arme Industrieprozesse

DLR Institut für Elektrifizierte Luftfahrtantriebe

ePHANT

EKI saxony

FBH Leibniz Ferdinand Braun Institut

FIBER FLOW

Fraunhofer IPMS

Fraunhofer IZM

Gantner instruments

iCampus COTTBUS

ihp

LAN COM East BERLIN

lausitz SCIENCE NETWORK

LSP Lausitz Science Park

Mesh4U

NM TECHNOLOGIES

OASYS

Perinet

Polytec

ROBERTA EDUCATION

SENTECH Erfolg durch Leistung

SpreeTec neXt Neue Fertigungstechnologien für Komponenten und Systeme der dezentralen Energietechnik

Thiem-Research GmbH WIR GEBEN MENSCHEN ZUKUNFT

Wirtschaftsförderung Brandenburg | WFBB

Kofinanziert von der Europäischen Union

Bundesministerium für Forschung, Technologie und Raumfahrt

Gold 9 qm

Silber 6 qm

Bronze 4 qm

1. Lausitz Science Park (LSP)

2. Actemium BEA

3. Wirtschaftsförderung Brandenburg (WFBB)

4. Roberta/Mesh 4U/NMI Technolologies

5. EKI saxony

6. Gantner Instruments

7. LAN-COM-East

8. Polytec

9. Perinet

10. SENTECH

11. BTU Cottbus-Senftenberg, Abteilung WTT

Sponsoren

INTERFACE SYSTEMS

ePHANT

FIBER FLOW

Aussteller

NM TECHNOLOGIES

Mesh4U

OASYS
Optoelektronische Sensoren für anwendungsnahe Systeme für Lebenswissenschaften und intelligente Fertigung

Perinet

Polytec

ACTEMIUM

Gantner instruments

Lausitz Science Park

DLR
Institut für Elektrifizierte
Luftfahrtantriebe

DLR
Institut für CO2-arme
Industrieprozesse

chesco
Center for Hybrid
Electric Systems
Cottbus

SENTECH
Erfolg durch Leistung

EKI
saxony

lausitz
SCIENCE
NETWORK

Fraunhofer
IZM

Fraunhofer
IPMS

b-tu
Brandenburgische
Technische Universität
Cottbus - Senftenberg

LAN
COM
East
BERLIN

FBH
Leibniz
Ferdinand
Braun
Institut

ihp
innovations
for high
performance
microelectronics

Thiem-Research GmbH
WIR GEBEN MENSCHEN ZUKUNFT

Wirtschaftsförderung
Brandenburg | WFBB
Kofinanziert von der
Europäischen Union

ROBERTA
EDUCATION

iCampus
COTTBUS

Sehr geehrte Damen und Herren,

erfolgreicher Strukturwandel braucht innovative Technologien, Wissenschaftler, die sie entwickeln und Unternehmen, die sie anwenden. Die iCampus Cottbus Conference (iCCC) 2026 bringt alle drei zusammen.

Sie bietet Schlüsseltechnologien wie Sensorik, Mikro-Elektro-Mechanischen Systemen und Künstlicher Intelligenz eine in Brandenburg einmalige Plattform. Ausgerichtet wird sie vom Innovationscampus Elektronik und Mikrosensorik Cottbus (iCampus), einem Projekt, das angetreten ist, die Lausitzer Wirtschaft zu digitalisieren.

Im iCampus kommt geballte Wissenschaftspower zusammen: Forscherinnen und Forscher der BTU Cottbus- Senftenberg arbeiten hier gemeinsam mit Kolleginnen und Kollegen namhafter Leibniz- und Fraunhofer-Institute und der Thiem Research GmbH.

Seit Ende 2019 entwickeln sie neuartige Sensorik und zugehörige Kommunikationstechnik für Industrie 4.0, Umwelt- und Agrarsensorik sowie Medizin und Life Sciences: Von Drohnenschwärmen zur Überwachung von Großanlagen oder Tagebaufolgelandschaften bis hin zur Datenübertragung aus Kuhmägen. In vielen Anwendungsfeldern wird bereits gemessen und kommuniziert, werten schlaue Systeme Daten zielgerichtet aus und können so Produkte und Prozesse verbessert und beschleunigt werden.

Die erste Konferenz vor zwei Jahren stand noch ganz im Zeichen des „Kennenlernens“: Akteure vernetzten sich, Arbeitsgebiete wurden vorgestellt. Die iCCC2026 geht einen Schritt weiter und zeigt, wie aus Forschung Anwendungen entstehen und aus Anwendungen Unternehmen.

Workshops zu Entrepreneurship und 5G, Live-Demonstrationen und junge Start-ups, die sich am Messestand präsentieren, setzen neue Akzente. Dass jede Session mit einem Impulsvortrag aus der Industrie beginnt, ist ein weiteres klares Signal: Ob LEAG oder Deutsche Bahn – hier ansässige Unternehmen formulieren ihre Bedarfe und unsere iCampus-Wissenschaftlerinnen und Wissenschaftler liefern Lösungen.

So entsteht Wertschöpfung, und so entstehen starke Partnerschaften für den Wissenschafts- und Wirtschaftsstandort Cottbus.

Ich wünsche allen Teilnehmerinnen und Teilnehmern einen erfolgreiche Konferenz, inspirierende Gespräche und viele neue Ideen!

Prof. Dr. p.h. habil. Gesine Grande
Präsidentin der BTU

Liebe Tagungsteilnehmerinnen und Teilnehmer,

im Namen des Bundesministeriums für Forschung, Technologie und Raumfahrt (BMFTR) begrüße ich Sie herzlich zum Innovationscampus Cottbus. Wir fördern den iCampus um die enge Vernetzung und den Transfer von anwendungsnaher Forschung in die regionale Wirtschaft zu unterstützen. Dies ist sowohl zentraler Baustein des Strukturstärkungsgesetzes als auch der neuen HighTech Agenda Deutschland (HTAD).

Die Bundesregierung definiert in der HTAD die Mikroelektronik als eine von sechs Schlüsseltechnologien. Im Oktober 2025 wurde deshalb die erste ressortübergreifende Mikroelektronik-Strategie für Deutschland verabschiedet. Kerngedanke: wir müssen die drei „Fs“ – Forschung, Fertigung und Fachkräfte – konsequent zusammendenken. Dies geschieht im iCampus: Entwicklung hochmoderner Sensorlösungen, rasche Nutzung durch lokale Unternehmen und gleichzeitig Schaffung von attraktiven Ausbildungs- und Beschäftigungsperspektiven.

Die konkreten Anwendungsbeispiele dieses Projektes zeigen die Leistungsfähigkeit und Vielseitigkeit moderne Mikrosensorik. Entwickelt werden z.B. miniaturisierte Sensoren zur Überwachung fluider Kraftstoffe als wichtiger Beitrag zu Sicherheit, Effizienz und Nachhaltigkeit. Für Gesundheitsanwendungen entstehen Radar- und Ultraschallsysteme zur berührungslosen Überwachung von Patientinnen und Patienten. Hinzu kommen In-situ-Sensoren zur Überwachung von Nutzvieh, die Tierwohl und Effizienz in der Landwirtschaft unterstützen. Hochfrequenz-Sensorknoten dienen als vernetzte Bausteine für intelligente Systeme in Industrie, Infrastruktur und Umwelt.

Die ICCC2026 als Transfer- und Netzwerkkonferenz ist ein wichtiger Baustein, um Wissenschaft, Industrie und Start-ups noch besser zu vernetzen und den Austausch über technologische Entwicklungen zu fördern und nachhaltig in der Region zu verankern. Nutzen Sie diese Plattform, um Wissen zu teilen, neue Kooperationsmöglichkeiten zu entdecken und gemeinsam die digitale Zukunft der Region zu gestalten. Ihr Engagement ist der Schlüssel zum Erfolg - lassen Sie uns gemeinsam Innovation, Beschäftigung und nachhaltige Strukturentwicklung vorantreiben.

Ich wünsche Ihnen inspirierende Vorträge, spannende Diskussionen und zahlreiche fruchtbare Partnerschaften.

Ministerialrat Dr.-Ing. Fabian Kohler
Referatsleiter BMFTR
Referat Mikroelektronik, Supercomputing, Robotik

Sehr geehrte Damen und Herren,

liebe Teilnehmerinnen und Teilnehmer der zweiten iCCC-Konferenz!

Die Lausitz hat in den vergangenen Jahrzehnten tiefgreifende Umbrüche erlebt – und sie hat daraus Stärke gewonnen. Wir haben es hier mit einer Region zu tun, die Transformation kann. Strukturwandel fordert Mut, Zusammenhalt und Zukunftsoptimismus – Transformation ist nichts für Feiglinge. Zugleich eröffnet er die einmalige Chance, die Geschichte dieser Region als Erfolgserzählung von Innovation, Wissen und Verantwortung fortzuschreiben.

Wissenschaft und Forschung sind dabei der Schlüssel. Mit der BTU Cottbus-Senftenberg, der neu entstehenden Medizinischen Universität Lausitz – Carl Thiem und dem Lausitz Science Park bündeln Land, Bund und regionale Partner ihre Kräfte, um ein Innovationsökosystem von internationaler Ausstrahlung aufzubauen. Hier verbinden sich exzellente Grundlagen- und angewandte Forschung, Technologietransfer und wirtschaftliche Dynamik. Denn in der Lausitz haben sich Wissenschaft und Forschung keine Elfenbeintürme gebaut – hier wird für den konkreten Nutzen gearbeitet, mit spürbaren Auswirkungen auf Wertschöpfung, Beschäftigung und die Attraktivität der Region für die dringend benötigten Fachkräfte. Sie wissen das Land an Ihrer Seite!

Der iCampµs ist eine tragende Säule dieses Ökosystems. Als Innovationscampus für Elektronik und Mikrosensorik bringt er Universität, außeruniversitäre Forschung, Unternehmen und Start-ups zusammen, um gemeinsam an Zukunftstechnologien zu arbeiten und diese in konkrete Anwendungen und Produkte vor Ort zu überführen. Sensorik, MEMS, Digitalisierung und Künstliche Intelligenz werden hier als Schlüsselfaktoren des Strukturwandels verstanden – anschlussfähig und praktisch nutzbar für Industrie 4.0, Energie, Umwelt und Gesundheit.

Das prall gefüllte Programm der iCampµs Cottbus Conference 2026 führt eindrucksvoll vor Augen, welche Expertise mittlerweile in Cottbus zuhause ist. Die Konferenz schafft Raum für Austausch, Vernetzung und neue Kooperationen – von der Bio- und Umweltsensorik über Kommunikations- und Energietechnologien bis hin zu Materialforschung, Gesundheit und Sport. Das große Interesse aus Wirtschaft, Forschung und Verwaltung zeigt, dass die Lausitz inzwischen als Labor und Schaufenster für gelingenden Strukturwandel wahrgenommen wird.

Der iCampµs verkörpert, was wissenschaftliche Exzellenz ausmacht, wenn sie sich regional verwurzelt: Neugier, Tatkraft und den Willen, Zukunft in eigener Verantwortung zu gestalten. So wird ein neues, hochinteressantes Kapitel in der Geschichte der Lausitz aufgeschlagen – ein Fortsetzungsroman, der zeigt, dass hier nicht nur Kohlegeschichte geschrieben wurde, sondern dass auch die neue Gegenwart und die vielversprechende Zukunft einer innovativen, solidarischen und lebenswerten Region viel zu erzählen haben.

Ich wünsche Ihnen spannende Vorträge, steile Lernkurven und inspirierende Gespräche!

Dr. Manja Schüle
Ministerin für Wissenschaft, Forschung und Kultur des Landes Brandenburg

Keynotes - Plenarvorträge

K1 **Digitale Resilienz: Warum Sensorik, KI und Robotik Europas Zukunft bestimmen** **18**
Rolf Härdi, Deutsche Bahn AG

K2 **Wenn Elektronen auf Photonen treffen: Neuartige heterogene Integrationstechnologien als Basis der nächsten Generation medizinischer und photonischer Systeme** **19**
Dr. Gabriel Kittler, X-Fab Erfurt

K3 **Optik und Photonik sind forschungsstarke Schlüsseltechnologien und werden es bleiben** **20**
Prof. Dr. Michael Totzeck, ZEISS Expert Ladder – Fellow

K4 **Sensors in Medical Imaging Applications – Challenges, Solutions and Opportunities** **22**
Dr.-Ing. Michael Hosemann, Siemens Healthineers

Invited Talks - Übersicht

I1 **Die Nadel im Heuhaufen finden: Datengetriebene Alterungserkennung technischer Assets durch unüberwachte Zeitreihenanalyse** **23**
Dr.-Ing. Konstantin Jonas, Deutsche Bahn AG

I2 **Multimodale Sensorik und KI in der Sportmedizin: Von der Leistungsdiagnostik zur supportiven Bewegungstherapie** **24**
Dr. phil. Barlo Hillen, Institut für Sportwissenschaft, Mainz

I3 **Product-Market: Fit or Die!** **26**
Dr. Andreas Olmes

I4 **Energie ist Leben** **28**
Dr. Thomas Porsche, LEAG

I5 **Multisensorik für die Prozessindustrie** **29**
Dr. Attila Michael Bilgic, Krohne

I6 **ECOSENSE – A smart DISTRIBUTED sensor Network for forests** **30**
Prof. Dr. Ulrike Wallrabe, University of Freiburg, IMTEK

I7 **Networked Sensors: The Last Expansion of the Internet** **32**
Dr.-Ing. Karsten Walther, Perinet GmbH, Berlin

I8 **Ultraschall-Gaszähler für die eichpflichtige Durchflussmessung von Gasen** **33**
Dr. Eric Starke, Endress+Hauser SICK, Ottendorf-Okrilla

I9 **4G/5G Netze mit Boldyn EVO für verbesserte Logistik** **34**
Rüdiger Hnyk, Boldyn Networks Deutschland

Session I - Condition Monitoring

Chair: Dr.-Ing. Konstantin Jonas, Deutsche Bahn AG

1.1 **Die Nadel im Heuhaufen finden: Datengetriebene Alterungserkennung technischer Assets durch unüberwachte Zeitreihenanalyse** **23**
Dr.-Ing. Konstantin Jonas, Deutsche Bahn AG

1.2 **Field Data optimized Lifetime Prediction of Li-Ion Batteries in EV Fleets** **36**
T. Lehmann, Fraunhofer IVI, Dresden

1.3 **Smart Sensor for predictive maintenance of a milling machine** **42**
J. Liebermann, M. Lehmann, D. Mayer, A. Schneider, A. B. Gowri, Fraunhofer IIS, Dresden

1.4 **Berstdruckvorhersage von faserverstärkten Druckbehältern anhand von Schallemissionen** **46**
E. D. Kästle, A. Ghaznavi, E. Duffner, BAM Bundesanstalt für Materialforschung und -prüfung, Berlin

Session II - Gesundheit und Sport

Chair: Dr. phil. Barlo Hillen, Institut für Sportwissenschaft, Universität Mainz

2.1 **Multimodale Sensorik und KI in der Sportmedizin: Von der Leistungsdiagnostik zur supportiven Bewegungstherapie** **24**
Dr. phil. Barlo Hillen, Institut für Sportwissenschaft, Mainz

2.2 **Integrative Bewertung von Mikrozirkulationsstörungen bei Post-Covid 19** **51**
L. Schlierkamp-Voosen, B. Hillen, V. Weber, A. Tomaskovic, D. T. Ochmann, S. Zentgraf, D. A. López, E. W. Neuberger, P. Simon, E.-V. Griemert, Universität Mainz, Mainz

2.3 **Mobility Monitoring for Geriatrics: Gait Detection using e-Textile** **55**
T. Steinmetzer, S. Michel, BTU Cottbus-Senftenberg, Cottbus

2.4 **Implantable MEMS Sensor Platform for Measuring Heart Activity** **59**
S. Schäfer, P. Detemple, O. Haverbeck, J. Magomajew, S. Schmitt, Fraunhofer IMM, Mainz, M. Haberbusch, F. Moscato, Medizinische Universität Wien, Wien (Austria)

2.5 **Trainingssteuerung im Leistungssport: Integration multimodaler Daten** **63**
A. Burmann, F. Lauf, Fraunhofer ISST, Dortmund, S. Schröder, Enari GmbH, Garching bei München

Session III - Workshop Entrepreneurship

Chair: Dr. Andreas Olmes

3.1 **Product-Market: Fit or Die!** **26**
Dr. Andreas Olmes

Session IV - Energiewirtschaft und Material & Prozesstechnologie

Chair: Dr. Thomas Porsche, LEAG

4.1 **Energie ist Leben** **28**
Dr. Thomas Porsche, LEAG

4.3 **Untersuchung der optischen Eigenschaften von Palladiumdünnschichten** **67**
P. Frenzel, R. Täschner, WHZ Westsächsische Hochschule Zwickau, Zwickau

4.4 **Voll digitalisiertes Reallabor Wasserstofftankstelle der BAM** **71**
R. Bock, C. Tiebe, F. Lückoff, G. Mair, BAM Bundesanstalt für Materialforschung und -prüfung, Berlin

4.5 **Portable NIR Spectroscopy for Reliable Characterization** **76**
M. Nobari, H. Engelke, I. Jabłoński, Fraunhofer IPMS, Cottbus, F. Eckert, M. Nobari, I. Jabłoński, BTU Cottbus-Senftenberg, Cottbus

4.6 **Halbleiter für Quantumcomptuting: Wertschöpfung und chemischer Fingerabdruck** **80**
O. C. Ernst, D. Uebel, R. Brendler, A. Handschick-Ernst, J. Acker, BTU Cottbus-Senftenberg, Senftenberg

4.7 **Mehr als Bionik - Biofunktionelle Polymere als innovative neue Materialien** **84**
S. Rosencrantz, J. Storsberg, R. R. Rosencrantz, Fraunhofer IAP, Potsdam und BTU Cottbus-Senftenberg, Senftenberg

Session V - Bio- und Umweltsensorik

Chair: Dr. Attila Michael, Bilgic, Krohne
Prof. Dr. Ulrike Wallrabe, University of Freiburg, IMTEK

5.1 **Multisensorik für die Prozessindustrie** **29**
Dr. Attila Michael Bilgic, Krohne

5.2 **ECOSENSE – A smart DISTRIBUTED sensor Network for forests** **30**
Prof. Dr. Ulrike Wallrabe, University of Freiburg, IMTEK

5.3 **Entwicklung von Plasmonischen On-Chip Sensoren** **87**
S. Reiter, A. Sengül, P.-G. Nitsch, F. Berkmann, I. A. Fischer, BTU Cottbus-Senftenberg, Cottbus,
C. A. Chavarin, J. Schlipf, C. Mai, C. Wenger, IHP Leibniz-Institut für innovative Mikroelektronik, Frankfurt (Oder)

5.4 **Auf dem Weg zur radarbasierten Rinderüberwachung** **90**
M. Ulbricht, V. Sark, R. T. Syed, J. P. Gose, IHP Leibniz-Institut für innovative Mikroelektronik, Frankfurt (Oder),
G. Hoffmann, ATB Leibniz-Institut für Agrartechnik und Bioökonomie, Potsdam

5.5 **Chemofluidische Schaltkreise – Plattform zur Automatisierung** **94**
A. Richter, D. Gruner, M. Shahadha, F. Obst, T. Spitzner, J. Dohndorf, J. Paul, TU Dresden, Dresden,
M. Menschikowski, M. Friedemann, Universitätsklinikum Carl Gustav Carus, Dresden, J. Paul,
Diamics GmbH, Dresden

Session VI - Kommunikation

6.1 **Networked Sensors: The Last Expansion of the Internet** **32**
Dr. Karsten Walther, Perinet GmbH, Berlin

6.2 **GigaCoupler – A scalable Opto-Isolator with Gbps-Data Transmission** **98**
K. Schmieder, T. Schneider, R. Kirrbach, Fraunhofer IPMS, Dresden

6.3 **Dezentrale, Mesh-basierte Plattform zur Überwachung von Nutztieren** **103**
M. Ferus-Comelo, F. Galka, D. Strauß, DHBW Stuttgart, Stuttgart

6.4 **RFID und BLE Tracking im Industriellen Umfeld** **107**
T. Krause, M. Richter, P. Näser, BTU Cottbus-Senftenberg, Cottbus

Session VII - Ultraschallsensorik und Mobilität

Chair: Dr. Eric Starke, Endress+Hauser SICK, Ottendorf-Okrilla

7.1 **Ultraschall-Gaszähler für die eichpflichtige Durchflussmessung von Gasen** **33**
Dr. Eric Starke, Endress+Hauser SICK, Ottendorf-Okrilla

7.2 **1D phased array for imaging applications in air utilizing MEMS** **111**
M. Jongmanns, J. M. Monsalve, Fraunhofer IPMS, Dresden

7.3 **Ultrasound-Based Interferometric Hip Implant Monitoring** **115**
J. Lützelberger, K. S. Drese, Hochschule Coburg, Coburg

7.4 **Modellierung einer Heat-Pipe-gestützten Thermalregulierung von HT-PEM-Brennstoffzellen in der Luftfahrt - Numerische Bestimmung geometriespezifischer Wärmeübergangskoeffizienten** **119**
F. Kramer, F. Franke, M. Kober, S. Kazula
Deutsches Zentrum für Luft- und Raumfahrt e.V. (DLR), Cottbus

Session VIII - Workshop 5G

Chair: Rüdiger Hnyk, Boldyn Networks, Deutschland

8.1 **4G/5G Netze mit Boldyn EVO für verbesserte Logistik** **34**
Rüdiger Hnyk, Boldyn Networks Deutschland

Poster Session

Akustik & Ultraschallsensorik

P1 **Analysis of Ultrasound Echoes for Object Properties Characterisation** **124**
P. Suawa, C. Herglotz, BTU Cottbus-Senftenberg, Cottbus, M. Jongmanns, Fraunhofer IPMS, Dresden

P2 **Physikalisch basiertes virtuelles Prototyping von Mikrosystemen** **128**
G. Schrag, TU München, München

P3 **Ultrasonic Waveguide for Multiphase Level Monitoring** **132**
M. Piechaczek, A. Backer, T. Roß, S. Tietze, K. S. Drese, Hochschule Coburg, Coburg

Bio- & Umweltsensorik

P4 **Contributions of Shifted Excitation Raman Difference Spectroscopy...** **136**
K. Sowoidnich, A. Müller, M. Maiwald, FBH Ferdinand-Braun-Institut, Berlin

P5 **Electrical Characterization of CeO 2 /SnO 2 Hydrogen Sensors** **139**
A. Mudundi, C. A. Chavarin, C. Wenger, IHP Leibniz-Institut für innovative Mikroelektronik, Frankfurt (Oder),
A. Kalra, I. A. Fischer, C. Morales, J. I. Flege, BTU Cottbus-Senftenberg, Cottbus

P6 **Nanostrukturierte Dünnschicht-Schottky-Fotodioden für SWIR** **143**
L. Shala, J. Knobbe, L. Augel, Fraunhofer IPMS, Cottbus, L. Shala, L. Augel, BTU Cottbus-Senftenberg, Cottbus

P7 **Ammonium and Potassium Ion-Sensitive Field-Effect Transistors** **147**
L. T. Kühne, C. Beale, M. Wambold, E. Kurth, O. R. Hild, Fraunhofer IPMS, Dresden,
C. Ruffert, Fraunhofer IPMS, Cottbus

P8 **IR Soil Moisture – Contactless Detection of Forest Soil Dryness...** **151**
U. Franke, 5micron GmbH, Berlin

P9 **Miniaturisiertes Quantenlichtmodul zur Mikroplastikdetektion** **155**
K. Bodenhagen, G. Blume, A. Sahm, F. Mauerhoff, S. Ramelow, K. Paschke,
FBH Ferdinand-Braun-Institut, Berlin, A. U. R. Sherwani, Humboldt-Universität zu Berlin, Berlin

P10 **Realisierung eines Demosystems für einen neuartigen Ansatz einer intelligenten Hyperspektralkamera** **159**
H. Engelke, M. Jongmanns, F. Eckert, Fraunhofer IPMS, Dresden

P11 **Miniaturisierte Spektralanalysesysteme im Kontext von Lebensmittelqualität** **162**
H. Engelke, T. Pügner, Fraunhofer IPMS, Dresden

P12 **Design and Integration of Functional Blocks for Sub-THz Radar Systems** **166**
B. Sutbas, V. Ertürk, R. Hasan, C. Carta, IHP Leibniz-Institut für innovative Mikroelektronik, Frankfurt (Oder),
P. Tschammer, V. Krozer, W. Heinrich, FBH Ferdinand-Braun-Institut, Berlin, U. Maaß, U. T. Yeasmin,
A. Gäbler, I. Ndip, Fraunhofer IZM, Cottbus

P13 **Development of Bolus Sensor for Cattle Rumen Monitoring** **170**
C. Brockmann, A. Gäbler, U. T. Yeasmin, S. C. Pasupuleti, P. Arnold, U. Maaß, I. Ndip,
Fraunhofer IZM, Cottbus

Condition Monitoring / Predictive Maintenance

P14 **Analyse industrieller Prozessdaten mit KI Agenten** **173**
K. Gehrke, S. Spiesecke, T. Bartel, FBH Ferdinand-Braun-Institut, Berlin

P15 **Untersuchungen der Schmierfettverteilung in Labyrinthdichtungen durch Kapazitätsmessung** **175**
A. Bürger, S. Simon, BTU Cottbus-Senftenberg, Senftenberg

P16 **KI-gestützte Messtechnik in der Siliziumwafer-Produktion** **179**
D. Kaufmann, Gantner Instruments GmbH, Lauf an der Pegnitz

P17 **Fire risk assessment using advanced sensing at the edge of drones** **184**
S. Devi, BTU Cottbus-Senftenberg, Cottbus, I. Jablonski, Fraunhofer IPMS, Dresden

P18 **Early Prediction of H2 Concentration and Sensor Response Using MLP** **188**
R. Sarif, C. Tiebe, BAM Bundesanstalt für Materialforschung und -prüfung, Berlin, C. Herglotz, BTU Cottbus-Senftenberg, Cottbus

P19 **Passive MEMS-Sensoren für das Condition Monitoring** **194**
P. Schmitt, M. Hoffmann, Ruhr-Universität Bochum, Bochum, P. Schmitt, mechIC GmbH, Bochum

Energiewirtschaft

P20 **Umsetzbarkeit eines autarken solarbetriebenen HiFi-Lautsprechersystems** **198**
A. Arnold, M. Eberle, T. Stoppe, F. Ellinger, TU Dresden, Dresden

P21 **Public Communication on the Energy Transition - NLP Approach** **202**
M. Petrosjan, D. Korrapati, C. Hoffmann, F. Müsgens, BTU Cottbus-Senftenberg, Cottbus

P22 **Modellierung atmosphärischer Turbulenz für Windlast-Simulationen** **208**
M. Klein, BTU Cottbus-Senftenberg, Cottbus

P23 **Auswirkungen von KI auf die Betriebskosten und das Abwärme Potential** **212**
G.-R. Schmidt, TU Berlin, Berlin, T. Kneiske, Fraunhofer IEG, Cottbus

P24 **Effects of Surface Functionalization on Convective Heat Transfer** **216**
S. R. G. Polasanapalli, M. Klein, H. Schmidt, BTU Cottbus-Senftenberg, Cottbus

Gesundheit & Sport

P25 **Mobile SERDS Device for in vivo Skin Antioxidant Monitoring** **220**
K. Sowoidnich, A. Müller, M. Maiwald, FBH Ferdinand-Braun-Institut, Berlin, A. Pfennig, S. K. Starke, T. König, M. Schmidt-Hieber, Medizinische Universität Lausitz – Carl Thiem, Cottbus

P26 **Digital unterstützte Betreuung von Pflegeeinrichtungen – Vorstellung des DUAB-Projektes** **223**
K.-P. Braun, R. Märker, A.-V. Klement, MVZ Dr. Braun GmbH, Cottbus

P27 **Mikrointegriertes Quantenlichtmodul für die hyperspektrale Bildgebung** **227**
N. Werner, S.-A. Seeger, F. Mauerhoff, A. Sahm, J. Zender, D. Feise, G. Blume, S. Ramelow, K. Paschke, FBH Ferdinand-Braun-Institut, Berlin, A. U. R. Sherwani, Humboldt-Universität zu Berlin, Berlin

P28 **Explainability in ConvNeXt V2 for Spectral Mammography** **231**
W. Ghorbel, A. Siyavashi, C. Herglotz, BTU Cottbus-Senftenberg, Cottbus

P29 **Effekte der Quantisierung auf RAG Language Modelle für Parkinson** **235**
P. Gaudl, C.-A. Holst, V. Lohweg, Technische Hochschule Ostwestfalen-Lippe, Lemgo

Kommunikation

P30 **Möbius-Shapley: Native Feature Attribution for Quantum Logic** **239**
S. Dhahbi, I. Schmitt, BTU Cottbus-Senftenberg, Cottbus

P31 **A Cybersecurity Approach for Visual Data Flow in Digital Twins** **244**
H. Zargariasl, A. Faraj, C. Borck, M. Behm, A. Arnouk, C. Herglotz, BTU Cottbus-Senftenberg, Cottbus

P32 **Visualizing Score Contributions in BBQ-Trees** **248**
A. Stahl, BTU Cottbus-Senftenberg, Cottbus

P33 **Securing Custom IEEE 802.11 Based Mesh Networks on ESP32 IoT Modules** **252**
P. Mykytyn, R. Rotta, BTU Cottbus-Senftenberg, Cottbus

P34 **A 5G power-hardware-in-the-loop testbed for smart power systems** **256**
R. C. Martínez, E. W. Zurita, J. Schiffer, N. R. Ramireddy, BTU Cottbus-Senftenberg, Cottbus, J. Schiffer, Fraunhofer IEG, Cottbus

P35 **Step by Step from Modbus to secure IoT** **260**
R. Karnapke, Perinet GmbH, Cottbus, K. Walther, Perinet GmbH, Berlin

P36 **Performance Assessment of 5G for Remote Sensing** **264**
K. Shahin, C. Herglotz, BTU Cottbus-Senftenberg, Cottbus

P37 **Mobility-Induced Sensitivity of UAV-based Nodes to Jamming** **268**
P. Mykytyn, O. Yener, P. Langendoerfer, BTU Cottbus-Senftenberg, Cottbus, R. Chitauro,
IHP Leibniz-Institut für innovative Mikroelektronik, Frankfurt (Oder)

P38 **Further Development of a Measurement Method assessing the Coverage Quality of a private Standalone 5G network** **272**
J. Levon, R. C. Martinez, K. Shahin, BTU Cottbus-Senftenberg, Cottbus, I. Alkhasli,
Fraunhofer IKTS, Dresden

P39 **Adaptive 5G Gateway for Smart Livestock Monitoring** **276**
N. P. Doe, H. Zagariasl, C. Herglotz, BTU Cottbus-Senftenberg, Cottbus

P40 **MooNet: Low-Power Mesh Backbone for Livestock Monitoring** **280**
N. S. Chatharajupalli, R. Rotta, J. Nolte, BTU Cottbus-Senftenberg, Cottbus

P41 **Schnelle Überprüfung miniaturisierter optischer Transceiver** **284**
Y. Xu, K. Schmieder, T. Schneider, R. Kirrbach, Fraunhofer IPMS, Dresden, R. Tutsch,
TU Braunschweig, Braunschweig

Material- & Prozesstechnologie

P42 **Formation and role of the interface region in metal oxide layers** **290**
C. Morales, R. Tschammer, D. Guttmann, K. Henkel, J. I. Flege, BTU Cottbus-Senftenberg, Cottbus,
C. Alvarado, C. Wenger, IHP Leibniz-Institut für innovative Mikroelektronik, Frankfurt (Oder)

P43 **Cleaning effect on surface activated aluminum to aluminum wafer bonding** **294**
S. Schulze, T. Voß, P. Krüger, R. Lukose, P. Kulse, M. Wietstruck,
IHP Leibniz-Institut für innovative Mikroelektronik, Frankfurt (Oder)

P44 **In-situ-Ätzkontrolle für vertikale GaN-basierte Leistungsbauelemente...** **297**
R.-S. Unger, F. Brunner, E. Brusaterra, O. Krüger, FBH Ferdinand-Braun-Institut, Berlin,
K. Haberland, LayTec AG, Berlin, S. Peters, Sentech Instruments GmbH, Berlin

P45 **Calculated electromigration wind force in molybdenum disilicide** **301**
J. Baldauf, CiS Forschungsinstitut für Mikrosensorik GmbH, Erfurt

P46 **Quanteninterferenzmodule für die optische Kohärenztomographie...** **303**
G. Blume, D. Feise, A. Sahm, N. Werner, J. Zender, S. Ramelow, K. Paschke,
FBH Ferdinand-Braun-Institut, Berlin

Mobilität

P47 **The Importance of Predicting Battery Degradation** **307**
V. P. Joshi, R. Cepeda-Gomez, D. Knieschke, DLR Deutsches Zentrum für Luft- und Raumfahrt, Cottbus

P48 **Lastanalyse zur Energieprognose einer universitären AC-Ladeinfrastruktur** **311**
T. Scheppan, T. Fischer, D. Lehmann, M. Schenk, K. Malekian, BTU Cottbus-Senftenberg, Cottbus

P49 **Großbrandtests mit Elektrofahrzeugen für den Transport gefährlicher Güter** **314**
H. Soderer, T. Gleim, N. Böttcher, R. Tschirschwitz, F. Wille, BAM Bundesanstalt für
Materialforschung und -prüfung, Berlin

P50 **Drone Usage in industrial fields** **318**
P. Werban, M. Nattke, P. Naeser, BTU Cottbus-Senftenberg, Cottbus

P51 **Accelerating the development of electric propulsion systems for aviation...** **322**
M. Peciak, H. Witte, J. Engelmann, CHESCO GmbH, Cottbus

DOI 10.5162/iCCC2026/K1

Keynote 1

Digitale Resilienz: Warum Sensorik, KI und Robotik Europas Zukunft bestimmen

Rolf Härdi, Deutsche Bahn AG

Zusammenfassung:

Die Digitalisierung transformiert industrielle Wertschöpfungsketten fundamental. Am Beispiel des Bahnsektors zeigt diese Keynote, wie moderne Sensorik, Condition Based Monitoring, Robotik und Künstliche Intelligenz Instandhaltung, Sicherheit und Effizienz neu definieren.

Echtzeitdaten ermöglichen vorausschauende Wartung, erhöhen Systemverfügbarkeit und leisten einen entscheidenden Beitrag zur Klimaresilienz durch Ressourceneffizienz, Energieoptimierung und robuste Infrastrukturen. Gleichzeitig entsteht eine neue Disruptionswelle durch humanoide Robotik – insbesondere aus China – mit potenziell tiefgreifenden Auswirkungen auf Produktionsmodelle, Arbeitsmärkte und geopolitische Abhängigkeiten.

Ausgehend von konkreten Praxisbeispielen wird der Blick auf die europäische Industrie erweitert und eine ganzheitliche Bewertung technologischer, wirtschaftlicher und klimatischer Chancen sowie strategischer Risiken vorgenommen.

DOI 10.5162/iCCC2026/K2

Keynote 2

Wenn Elektronen auf Photonen treffen: Neuartige heterogene Integrations-technologien als Basis der nächsten Generation medizinischer und photonischer Systeme

Dr. Gabriel Kittler,
X-FAB MEMS Foundry GmbH

Zusammenfassung:

Die Einführung neuartiger heterogener Integrationstechnologien zur Kombination verschiedener Materialsysteme eröffnet Möglichkeiten, Innovationen voranzutreiben und neue Anwendungen zu entwickeln, um gesellschaftlichen Herausforderungen zu bewältigen. Die Konvergenz zwischen Elektronik und Photonik ist dabei eine Schlüsseltechnologie und erfordert die Schaffung neuer Ökosysteme und Modelle für die Zusammenarbeit. Im Zentrum steht die Evolution von diskreten MEMS hin zu hochintegrierten Systemen und heterogener Integration neuartiger Materialien und Chiplets (z. B. InP, LNO, 2D-Materialien, …) als Basis dafür. Diese Entwicklung ermöglicht neue Anwendungen in Automotive, Industrie, und Medizin, wie sie an Beispielen vorgestellt wird.

DOI 10.5162/iCCC2026/K3

Keynote 3

Optik und Photonik sind forschungsstarke Schlüsseltechnologien und werden es bleiben

Prof. Dr. Michael Totzeck,
ZEISS Expert Ladder – Fellow

Zusammenfassung:

Optik und Photonik sind Schlüsseltechnologien mit einer extrem großen Breitenwirkung. Die Mehrzahl der aktuellen Hot-spots der Innovationsaktivitäten haben ein starkes optisches Standbein. Digitalisierung und KI wären ohne die optische Lithografie als Enabler der Mikroelektronik nicht möglich. Gleichzeitig revolutionieren sie durch Technologien wie Computational Imaging und KI basierte Datenanalyse die Optik von Grund auf, indem sie neue Funktionen ermöglichen und damit neue Anwendungsbereich erschließen.

Doch nicht nur im Softwarebereich, auch in der optischen und phonischen Hardware spielt die Mikroelektronik eine entscheidende Rolle für die Optik: Mikroelektronische Produktionsverfahren ermöglichen neue Komponenten wie photonische integrierte Schaltkreise, variable Optiken, holografische Elemente und Metamaterialien. Damit werden optische Systeme auf mehreren Ebenen miniaturisiert. Das wiederum ermöglicht die Funktionsintegration für Anwendungen wie LIDAR, Smartphone-Sensoren, Mixed Reality, Advanced Manufacturing und andere.

Auch in der Quantentechnologie werden viele Anwendungen erst durch die Optik und Photonik ermöglicht. Sei es in der Quantenkommunikation mit verschränkten Photonen, dem Quantum Sensing mit NV Zentren oder dem Quantum Computing mit Atomen. Gleichzeitig wird die Nutzung von Quanteneffekten für die hochempfindliche Sensorik und Bildgebung erforscht.

DOI 10.5162/iCCC2026/K3

Die obigen Beispiele demonstrieren: Optik und Photonik sind nicht "ausgeforscht" sondern der Kernbestandteil eines komplexen Wertschöpfungsnetzes mit einer intensiven Innovationsaktivität. Dieses wiederum basiert auf einem lebendigen Ökosystem aus Forschung, Entwicklung und Applikation. Wegen ihrer Bedeutung sollten Optik und Photonik darin als eigenständiger Player wahrgenommen werden, denn es gibt keinen Grund anzunehmen, dass dieser Feedback-Kreislauf beendet ist. Ganz im Gegenteil.

DOI 10.5162/iCCC2026/K4

Keynote 4

Sensors in Medical Imaging Applications – Challenges, Solutions and Opportunities

Dr. Michael Hosemann,
Siemens Healthineers

Abstract:

In a world with a growing and aging population, medical imaging is a rapidly expanding application spanning a wide range of modalities and technologies. The talk will introduce imaging technologies and their sensors. Based on Siemens Healthineers' latest Alpha-class photon-counting computed tomography scanners, challenges and solutions for sensor development will be presented. Furthermore, opportunities for transfer into other fields will be highlighted.

DOI 10.5162/iCCC2026/I1

Invited Talk 1

Die Nadel im Heuhaufen finden: Datengetriebene Alterungserkennung technischer Assets durch unüberwachte Zeitreihenanalyse

Dr.-Ing. Konstantin Jonas,
Deutsche Bahn AG

Zusammenfassung:

Die Deutsche Bahn steht angesichts vielfältiger Herausforderungen rund um Zuverlässigkeit und Pünktlichkeit unter großem Handlungsdruck. Ursachen sind unter anderem technische Störungen, eine stark beanspruchte Infrastruktur und ein hoher Modernisierungs- sowie Instandhaltungsbedarf. Ungeplante Ausfälle technischer Assets sind dabei ein entscheidender Faktor. Sie ließen sich oft zuverlässig vorhersagen, wenn denn die richtigen Sensordaten in ausreichender Granularität vorliegen würden. In der Praxis zeigt sich jedoch: Viele Analyse-Use-Cases scheitern, weil entweder die passenden Sensoren fehlen oder die bestehende Datenbasis nicht ausreicht. Häufig werden so bereits viele Stunden in Daten-Vorverarbeitung, Analyse und Hypothesenbildung investiert, bevor klar wird, dass der gewählte Ansatz ins Leere läuft.

Im Projekt Deflector erproben wir deshalb einen Ansatz, der den klassischen Analyseprozess umkehrt: Unüberwachte Methoden erlauben es uns, technische Hypothesen automatisiert und direkt aus den verfügbaren Zeitreihendaten abzuleiten - ganz ohne vorher festgelegte Use-Cases. Auf diese Weise identifizieren wir datengetrieben frühzeitig potenzielle Alterungseffekte und fokussieren Analyseaufwand dort, wo die Daten tatsächlich belastbare Erkenntnisse ermöglichen.

Im Vortrag stellen wir diesen Ansatz vor und zeigen anhand typischer Anwendungsfälle, wie datengetriebene Innovationen effiziente und praxistaugliche Lösungen für die Deutsche Bahn ermöglichen.

DOI 10.5162/iCCC2026/I2

Invited Talk 2

Multimodale Sensorik und KI in der Sportmedizin: Von der Leistungsdiagnostik zur supportiven Bewegungstherapie

Dr. phil. Barlo Hillen,
Johannes Gutenberg-Universität, Mainz

Zusammenfassung:

Regelmäßige, individuell dosierte körperliche Aktivität zählt zu den wirksamsten Maßnahmen zur Prävention chronischer innerer Erkrankungen und ist ein zentraler Bestandteil moderner, evidenzbasierter Gesundheitskonzepte. Darüber hinaus spielt Bewegungstherapie eine entscheidende Rolle in der supportiven Behandlung von Menschen mit inneren Erkrankungen.

Demgegenüber stehen im Leistungs- und Spitzensport hohe Trainingsumfänge sowie Wettkampfanforderungen, die häufig unter herausfordernden Umwelt- und Belastungsbedingungen stattfinden. Ziel ist es, maximale Leistungsfähigkeit zu ermöglichen, ohne ein erhöhtes Risiko für Überlastung, Verletzungen oder gesundheitliche Langzeitfolgen einzugehen.

Die präzise Bestimmung der individuellen Beanspruchung sowie ein datenbasiertes Belastungs- und Regenerationsmanagement stellen dabei eine zentrale wissenschaftliche und technologische Herausforderung dar. Die Session „Sport & Gesundheit“ der iCCC 2026 adressiert diesen Spannungsbogen zwischen Prävention, Therapie und Leistungsoptimierung und zeigt anhand ausgewählter Beispiele das transformative Potenzial moderner mikrosensorischer, multimodaler und KI-gestützter Applikationen.

Die eröffnende Keynote stellt zwei interdisziplinäre Forschungsprojekte aus Sportmedizin, Prävention und Rehabilitation vor, die das Innovationspotenzial multimodaler Sensorik exemplarisch aufzeigen.

DOI 10.5162/iCCC2026/I2

Präsentiert werden ein synchronisiertes multimodales Messsystem mit KI-gestützter Infrarot-Thermografie in der kardiopulmonalen Leistungsdiagnostik sowie ein digitales, sensorbasiertes Betreuungskonzept zur supportiven Bewegungstherapie bei Menschen mit inneren Erkrankungen. Im Fokus steht dabei die Quantifizierung akuter und langfristiger physiologischer Anpassungsreaktionen des menschlichen Körpers unter körperlicher Belastung in den unterschiedlichen Umgebungs- und Anwendungsszenarien.

Die aufgeführten Inhalte verdeutlichen damit eindrucksvoll, wie multimodale Sensorik und künstliche Intelligenz nicht nur die sportmedizinische Leistungsdiagnostik, sondern auch die digitale Prävention und Therapie transformieren.

DOI 10.5162/iCCC2026/I3

Invited Talk 3

Product-Market: Fit or Die!

Dr. Andreas Olmes

Abstract:

Why do most start-ups fail — even when the technology is impressive, the team is smart, and funding is available?

The uncomfortable truth: not because of technology, but because of missing Product-Market Fit (PMF).

This workshop is not about buzzwords, polished pitch decks, or theoretical business plans. It focuses on the one question that truly determines success or failure: Does anyone genuinely need what you are building — and are they willing to pay for it?

Drawing on 18 years of venture capital experience, first-hand founder and CEO roles, and the analysis of hundreds of failed and successful start-ups, this session breaks down

- why investors say no to technically outstanding products,
- why many teams jump into “sales mode” while still misunderstanding the real customer problem,
- and how Product-Market Fit can be felt long before revenues, KPIs, or growth curves make it obvious.

Using real-world case studies — from spectacular failures to successful exits — participants learn how to identify true customer pain points, formulate and test value

DOI 10.5162/iCCC2026/I3

hypotheses, recognize misleading early signals, and understand why focusing on one single customer can be more powerful than chasing many.

The workshop combines Venture Capital perspective, founder reality, and practical tools: customer discovery, PMF indicators, pivot decisions, pricing signals, and the critical difference between a “nice-to-have” feature and a real painkiller.

If you don’t just want to build something, but want to build something that truly matters and scales, this workshop is for you.

Because: Life is too short to build something nobody wants.

DOI 10.5162/iCCC2026/I4

Invited Talk 4

Energie ist Leben

Dr. Thomas Porsche, LEAG

Zusammenfassung:

Die industrielle Entwicklung begann mit der Nutzung neuer Energiequellen. Durch Kohle wurden Dampfmaschinen angetrieben, die Elektrizität ermöglichte emissionsfreie Energie, Öl und Gas den unabhängigen Transport großer Mengen über weite Distanzen und Erneuerbare Energien eine moderne und nachhaltige Industrie.

Energie ermöglicht eine höhere Produktion, schnellere Herstellung, niedrigere Kosten, mehr Wohlstand und wird Schritt für Schritt nachhaltiger.

Dabei waren und sind die Energiesystem einem ständigen Wandel unterworfen und es steht die Frage, welche Anforderungen werden an das Energiesystem der Zukunft gestellt. Und – welche Aufgaben ergeben sich daraus für Wissenschaft und Technik.

Aus dem Energiesystem von der Erzeugung zum Verbrauch wird eine System von zentralen und dezentralen Energiekreisläufen mit Integration unterschiedlicher Akteure bis hin zur Sektorkopplung. Dies erfordert neben neuen Technologien ein neues, grenzüberschreitendes Energiemanagementsystem hoher Komplexität. Diese Aufgabe kann zukünftige durch den Einsatz komplexer Modelle (KI), die Daten aus neuen Materialien und Tools generieren und Entscheidungen vorbereiten bzw. treffen gelöst werden.

Beispiele sind virtuelle Kraftwerke, Drohnen im Tagebau, neue Sensoren für Condition Monitoring und Predictive Maintenance.

DOI 10.5162/iCCC2026/I5

Invited Talk 5

Dr. Attila Bilgic, Krohne

Multisensorik für die Prozessindustrie

Zusammenfassung:

Die Bewertung von Technologien hängt von den ins Auge gefassten Anwendungen ab. In der Prozessindustrie gilt: Jede einzelne Messstelle ist teuer, sowohl in der Schaffung und als auch in der Wartung. Für sie ist ein Prozessanschluss zu schaffen und sie ist mit Energie zu versorgen. Außerdem müssen die von der Sensorik erfassten Daten in geeigneter Form der Analyse zugänglich gemacht werden. Für einen kostengünstigen und effizienten Betrieb einer Anlage ist es daher wichtig, möglichst das Maximum aus einer Messstelle herauszuholen. Hier helfen Multisensorik, eine intelligente Vorverarbeitung, und ausreichend Konnektivität deutlich weiter.

In allen drei Bereichen sind in den letzten Jahren erhebliche Fortschritte erzielt worden: Einfache Sensorik wird immer günstiger, die Prozessoren, die in Messgeräten verbaut werden, werden - durch die Smartphoneentwicklung getrieben - immer leistungsstärker, und es gibt inzwischen Möglichkeiten, auch unter den spezifischen Anforderungen der Prozessindustrie das Internet in die Anlage zu bringen. Innovationssprünge sind möglich, wenn diese Enabler intelligent genutzt und Fehler vermieden werden. Natürlich spielt hier auch der smarte Einsatz von KI eine wichtige Rolle.

DOI 10.5162/iCCC2026/I6

Invited Talk 6

ECOSENSE – A SMART DISTRIBUTED SENSOR NETWORK FOR FORESTS

Prof. Dr. Ulrike Wallrabe,
University of Freiburg (IMTEK), Freiburg

Abstract:

Global climate change threatens ecosystem functioning worldwide. Forest ecosystems are particularly important for carbon sequestration. However, recurrent stresses, such as heat waves, floods, and droughts, increasingly endanger forests worldwide. Many relevant processes occur at small scales and high spatial heterogeneity and their interactions can be key players to amplify or dampen a system's response to stress. Currently, we are lacking the appropriate measuring, data and modelling tools allowing for comprehensive, real time quantification of relevant processes at high spatio-temporal coverage.

Consequently, the interdisciplinary research project ECOSENSE develops, implements, and evaluates a versatile, distributed, autonomous, intelligent sensor network based on novel microsensors tailored to the specific needs in harsh forest environments. All measurements need to happen in a minimally invasive manner in order not to disturb the ecosystem we are exploring. This requires consideration of size, weight, inertness, cleanliness, and robustness. The complex environment of a remote naturally structured forests with rain, wind, insects and mammals thereby poses its own challenges that the ECOSENSE sensors are facing.

The network comprises miniaturized sensors for 1) the quantification of gas and water fluxes, specifically isotope discriminated CO2 and volatile organic compounds (VOC, here isoprene) at ppb concentrations, and 2) for the detection of typical stress markers, such as active, laser induced chlorophyll fluorescence. ECOSENSE explores these distributed sensors from leaf level including low power electronics, to medium sized laser spectrometers, up to drones and eddy covariance towers.

DOI 10.5162/iCCC2026/I6

In our field site, the ECOSENSE Forest, we are currently operating 4500 continuous data streams, partially with a temporal resolution of a few seconds. Measured data is transferred in real-time to a sophisticated data base and can be explored for process analysis and AI enhanced simulation models providing the basis for future predictions of process-based alterations in ecosystem functioning and sustainability.

DOI 10.5162/iCCC2026/I7

Invited Talk 7

Networked Sensors: The Last Expansion of the Internet

Dr.-Ing. Karsten Walther,
Perinet GmbH, Berlin

Abstract:

Over the past few decades, we have witnessed the growth of the Internet. It began with mainframe servers but gradually expanded to smaller devices like PCs, laptops, smartphones, and edge computers. The latter have had a significant impact on how typical installations in industries and buildings are managed at an operational level. This success has always relied on easier communication between devices, while central IT systems maintained control. With the rise of AI, the advantages of this easier communication have become even more pronounced. Moreover, even tiny systems like sensors can now possess the computational power that standard PCs had just a few years ago, thanks to Moore's Law.

Now, we can create tiny but powerful networked sensors and actuators, effectively closing the last gap between the Internet and the physical world. First and foremost, you no longer need a translator (gateway) between the top level (IT) and the field level (field buses, analog interfaces, etc.), which simplifies the overall system, boosts performance, and allows for more secure operations than ever before. These tiny networked devices have become intelligent communication partners; they can provide you with information rather than just raw data and can perform tasks autonomously. Let's see how the world is changing this time with the expansion of the Internet?

DOI 10.5162/iCCC2026/I8

Invited Talk 8

Ultraschall-Gaszähler für die eichpflichtige Durchflussmessung von Gasen

Dr. Eric Starke,
Endress+Hauser SICK, Ottendorf-Okrilla

Zusammenfassung:

Gaszähler basierend auf dem Ultraschall-Differenzlaufzeitverfahren sind etablierter Standard sowohl in der Prozessmesstechnik als auch in der eichpflichtigen Abrechnungsmessung. Ihre wesentlichen Vorteile sind der geringe Einfluss auf die Strömung sowie die hohe Messgenauigkeit selbst unter anspruchsvollen Prozessbedingungen, wie z. B. hohem Druck, extremen Temperatur oder bei Verschmutzungen.

Ein wichtiger Baustein des Messsystems sind dabei die Ultraschallwandler, die für den zuverlässigen Betrieb unter den genannten Bedingungen speziell ausgelegt werden. Verschiedene Wandlerdesigns stehen zur Verfügung, um die vielfältigen Anforderungen unterschiedlicher Applikationen abzudecken.

Der Beitrag gibt zunächst eine Einführung in das Funktionsprinzip und die typischen Einsatzfelder von Ultraschallgaszählern in der eichpflichtigen Durchflussmessung. Anschließend werden die Herausforderungen bei der Auslegung und Implementierung der Ultraschallwandler erläutert und technische Lösungsansätze vorgestellt.

DOI 10.5162/iCCC2026/I9

Invited Talk 9

Rüdiger Hnyk, Boldyn, Networks Deutschland

Schnelle Fernlenkung in private 5G Netzen: Boldyn EVO verbessert mission-critical Logistik durch Ultra-Low-Latency Video

Zusammenfassung:

Boldyn Networks (Boldyn), weltweit führender Anbieter von gemeinsam genutzten Netzwerkinfrastrukturen und privaten Netzwerken, stellt mit schneller 4G/5G Fernlenkung eine Verbesserung der mission-critical Logistik durch Boldyn eigene Ultra-Low-Latency Video Technologie zur Verfügung.

Boldyn EVO (Edge Video Orchestration) verbessert durch Ultra-Low-Latency Video Übertragung die Reaktionsgeschwindigkeit der entfernt sitzenden Fahrer der AGV (Autonomous Guided Vehicle) LKW. Die AGV-LKW stellen ein Kernstück der internen Logistik für den Transport von Gütern von/zu den Eisenbahn-Terminals und den Verarbeitungszentren auf dem Werksgelände her.

Die 4G/5G Netzwerk optimierte Ultra-Low-Latency -Videoübertragung erlaubt es, Steuerbefehle der Fahrer schneller an die AGV-LKW zu übertragen. Dadurch lassen sich schnellere Reaktionszeiten erzielen und der Transport der Güter kann sicherer erfolgen bei gleichzeitig schneller Transport-Möglichkeit.

Die extrem niedrigen Ultra-Low-Latency Ende-zu-Ende Video Latenzen (Verzögerungen) von 0,1 Sekunden werden mit gleichzeitig praxistauglichen Bandbreiten für Full-HD Videos im 4G/5G Netz übertragen. Die Verzögerung liegt dabei deutlich unter dem gesetzlich geforderten Grenzwert von 0,2 Sekunden („Fernlenkungs-Gesetz“ BGBl. 2025 I Nr.176 vom 25.07.2025) und erfüllt damit die Voraussetzungen für sicherheitskritische Anwendungen in der Logistik.

DOI 10.5162/iCCC2026/I9

4G/5G Netze mit Boldyn EVO für verbesserte Logistik

“Die Tests zeigen, dass Echtzeit-Videoübertragung mit extrem niedriger Latenz unter realen Betriebsbedingungen heute bereits zuverlässig möglich ist. Dies ebnet den Weg zur Automatisierung von Prozessen in der Logistik. Voraussetzung dafür sind leistungsfähige 4G/5G Kommunikationsinfrastrukturen mit innovativer EVO Video Technologie, wie sie Boldyn bereitstellt.“, sagt Rüdiger Hnyk, Chief Operating Officer von Boldyn Networks Deutschland.

Globaler Marktführer mit starken Referenzen

Boldyn Networks ist der bevorzugte Konnektivitätspartner für Kunden aus verschiedenen Branchen, darunter Verkehrssysteme (Deutsche Bahn, London Underground, New York Subway), Luftfahrt (Flughafen Helsinki), Seeverkehr (Hafen Kokkola und OPCSA-Hafen in Finnland) und Gesundheitswesen (Universitätsklinikum Oulu, Finnland). Das Unternehmen unterstützt zudem Smart-City-Initiativen wie „Roma5G” in Rom und „Sunderland5G” in Großbritannien und stellt vor allem in Nordamerika Konnektivität für Stadien und Universitäten bereit.

DOI 10.5162/iCCC2026/1.2

Field Data optimized Lifetime Prediction of Li-Ion Batteries in Electric Vehicle Fleets

Thomas Lehmann, Fraunhofer Institut für Verkehrs- und Infrastruktursysteme, Dresden, Deutschland
thomas.lehmann@ivi.fraunhofer.de

Kurzfassung

Lithium-Ionen-Batterien (Li-Ion) bilden die Grundlage moderner Energiespeichersysteme, insbesondere auch in Elektrofahrzeugen (EV). Während Labortests eine gezielte Charakterisierung des Batterieverhaltens ermöglichen, bietet die Analyse von Felddaten eine Bewertung des Verhaltens in der Praxis. Beispielsweise können Degradationsmuster und die Betriebseffizienz unter verschiedenen Fahrbedingungen bestimmt werden. Diese Studie untersucht Methoden zur Erfassung, Verarbeitung und Interpretation von Felddaten von Li-Ionen-Batterien, die in EV-Flotten eingesetzt werden. Der Schwerpunkt liegt auf dem Vergleich von Methoden zur Vorhersage des Gesundheitszustands (State-of-Health, SOH). Diese reichen von einfachen Ladungszyklen-Zählungen bis hin zu Algorithmen des maschinellen Lernens, die die während des Betriebs beobachteten Degradationseffekte berücksichtigen. Diese Studie basiert auf der Auswertung von 44 Elektrobussen, die in mehreren Städten im Einsatz sind. Dies ermöglicht die Bewertung unterschiedlicher Operationsbedingungen. Die Ergebnisse unterstreichen die Bedeutung von Felddaten für die Verbesserung der SOH-Schätzung. Damit geht eine Erhöhung der Zuverlässigkeit, Sicherheit und Leistung von Batterien in Elektrofahrzeugflotten einher. Dies ebnet auch den Weg für intelligentere Flottenmanagementsysteme mit optimierten Routenplanungs- und Ladestrategien.

Abstract

Lithium-ion batteries (Li-ion) are the basis of modern energy storage systems, particularly in the context of electric vehicle (EV) fleets. While laboratory testing provides controlled insights into battery characteristics, field data analysis offers a critical perspective on real-world behaviour, degradation patterns, and operational efficiency under diverse driving conditions. This study explores methodologies for collecting, processing, and interpreting field data from Li-ion batteries deployed across EV fleets. Key focus area is the comparison of state-of-health (SOH) prediction methods. Reaching from simple cycle counting to machine learning algorithms that exploit the degradation effects observed during operation. This study is based on a set of 44 electric buses, operating in multiple cities. Hence different operation and charging regimes are involved. The findings underscore the importance of field data in enhancing SOH estimation and thus battery reliability, safety, and performance within EV fleets. This also paves the way for smarter fleet management systems using optimized route planning and charge strategies.

1 Introduction

Whitin this study the benefits of ageing prediction using field data analysis shall be presented. In the first section the state of the art is summarized. The used bus fleet and the generated dataset is described in the following section. Thereafter applied algorithms, ranging from basic Ah-counting based prediction methods to more advanced methods based on machine learning approaches, are presented. Finally, the benchmark results are discussed.

1.1 Ageing Prediction Methods

Lithium-ion battery ageing prediction methods span a diverse range of approaches, each tailored to capture the complex mechanisms that govern battery degradation over time. An overview of ageing degradation mechanism and their effects is given in (1).These methods are critical for optimizing battery performance, ensuring safety, and extending lifespan. The approaches can be roughly classified into the following categories:

- **Empirical Models**: These rely on extensive experimental data to establish degradation trends. By fitting mathematical functions to observed capacity fade or internal resistance growth over time, empirical models offer straightforward predictions. However, they often lack generalizability across different chemistries, operating conditions, or usage profiles.
  Common shape functions include exponential, polynomial, or Arrhenius-based models that correlate ageing with temperature, depth of discharge, and charge/discharge rates (2). Prediction frameworks often separate degradation into two categories: calendar ageing, which occurs due to chemical instability over time even when the battery is idle, and cycle ageing, which results from repeated charge-discharge cycles. Accurate models must account for both, as their relative impact varies by application. For instance, electric vehicles experience dominant cycle ageing, while battery stationary storage systems (BESS) may suffer more from calendar ageing.
- **Electrochemical Models**: These physics-based models delve into the internal workings of the battery, simulating processes such as solid electrolyte interphase

(SEI) layer growth, lithium plating, and electrolyte decomposition (3) (4). They provide high-fidelity insights into ageing mechanisms by solving coupled differential equations that represent mass transport, charge transfer, and thermodynamics. While computationally intensive, they are invaluable for understanding root causes of degradation and for designing batteries with improved longevity.

- **Machine Learning Models**: Leveraging large datasets from battery testing and real-world usage, machine learning approaches—such as neural networks, support vector machines, and random forests—can uncover complex, nonlinear relationships between input features (e.g., temperature, voltage, current) and ageing outcomes. These models excel in adaptability and scalability. However, they require careful validation and may struggle with extrapolation beyond the training data.
- **Hybrid Approaches**: To balance accuracy and computational efficiency, hybrid models integrate empirical or machine learning techniques with electrochemical insights. For example, a model might use physics-based equations to simulate SEI growth while employing data-driven corrections for temperature effects. These approaches aim to capture the strengths of each method, offering robust predictions across diverse operating conditions (5).

While it has been shown that SOH prediction can benefit from advanced methodologies as well as from detailed study of variety of ageing effects, in practise basic empirical models are applied. This is mainly due to the challenge of extrapolating the knowledge to new cell chemistries and to real world operation. Further challenges are the data acquisition and processing. These issues are addressed in (6) and (7).
Most research institutes lack a data basis that is suitable to apply advanced machine learning approaches. Instead, often operation cycles are tested in accelerate ageing tests.
Hence most ageing prediction methods are still exclusively based on physical considerations and laboratory tests. An overview of experimental ageing investigation is given in (8).

1.2 Field Data Analysis

As outlined in the previous subsection, field data analysis emerges as a promising approach to obtain valuable information about ageing mechanisms. The necessity of laboratory-based ageing tests may be diminished while maintaining a high degree of prediction accuracy.

During operation of batteries in electric vehicles or within BESS a Battery Management System (BMS) is collecting and checking sensor data. These data include current and voltage on cell level as well as temperature of selected cells. Also, SOH estimation is performed. This includes multiple parameters that describe the performance of the battery. Furthermore, mandatory key parameters have to be stored respectively updated during battery operation in order to keep track of supply chains, resource efficiency and performance values. This is dictated by the European Union regulation Battery Passport (9), which introduces a summary of regulations to improve second life applications and recycling.
Due to computational and memory constraints only undemanding algorithms may be used on the BMS. Thus, sensor data is often also streamed to a cloud server to allow a central monitoring as well as more complex SOH estimation algorithms. Furthermore, this allows comparison between different batteries of the same type, to evaluate variations introduced by different usage conditions.

In (10) the general data management, -transfer and analysis approach as well as the opportunities and challenges are described. They distinguish between different layers that provide information to operators and manufacturers.
Despite the great potential described e.g. in (11), there is very few public available research on field data analysis. This is due restrictions carried out by operating companies and manufacturers, that gather the data. One exception is (12), where the lifetime prediction from field data is discussed. The work focuses on the combination of laboratory experiments with field data analysis. Nevertheless, they emphasize the lack of availability of large publicly accessible datasets, including validation data under real-world operating conditions.

2 Benchmark Data

A fleet of electric buses was monitored from mid-2022. The relevant signals and their time resolutions are listed in Table 1. Beside battery signals also vehicle distance and speed are used for identifying data gaps.

Signal	Unit	Time resolution
Current	A	100 ms
Voltage	V	100 ms
Temperature	°C	100 ms
state of charge (SOC)	%	On value-change
mileage	km	1 s
speed	km/h	100 ms

***Table 1**: Measured signals and time resolution.*

In addition to analysing the data during normal operation dedicated capacity checkup tests were performed. These are further described in Section 2.2. A coarse description of the operation data is given in the next subsection. A detailed analysis, including an energy consumption estimation can be found in (13).

2.1 Operation Data

The buses typically operate 18h a day and use opportunity charging. This means that they are fast charged at designated stations. However, this intermediate charging is only partially sufficient to provide the necessary energy.

Therefore, they are also fully charged overnight in the depot. A short section of the current and voltage data is shown in Figure 1. The total time sequence is divided into different sections and classified into categories describing the operation state. For each section key parameter like charge throughput or current rate histograms are calculated. These values can be used to find suitable training data for machine learning algorithms or to determine the operation profile. This step also includes the detection of data gaps and anomalies, which is of importance for accurate SOH estimation. Flawed data could otherwise compromise the subsequent predictions.

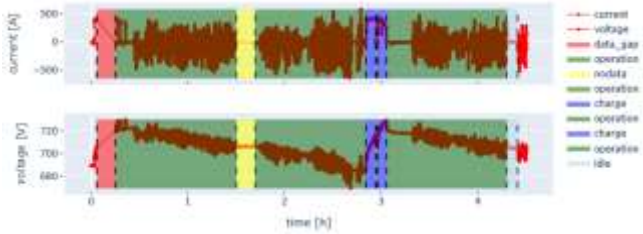


***Figure 1**: voltage and current signal during typical operation. Segments are grouped into different categories to simplify subsequent analysis*

2.2 Reference Tests

In order to validate estimations reference capacity tests are carried out on specified buses from time to time. This is very effortful, since buses have to be taken out of normal operation and the test is not fully automated. Furthermore, the procedure is very time consuming. One checkup test, which is described below lasts at least 24h. The test consists of the following steps:

- bus is driven until BMS-SOC reaches nearly zero percent
- if possible, the battery is further discharged (e.g. using heating)
- idle phase of at least 30 minutes is applied for battery relaxation
- charging with small current rate below $\frac{1}{10}C$ until BMS-SOC reaches 100%
- eventually further charging with constant voltage (CV) or pulsed charging

In Figure 2 a selection of these checkup measurements are compared with the open circuit voltage (OCV) curve. Reduced capacity leads to a compression of the charge curves. Hence cutoff voltage is reached earlier during charging.

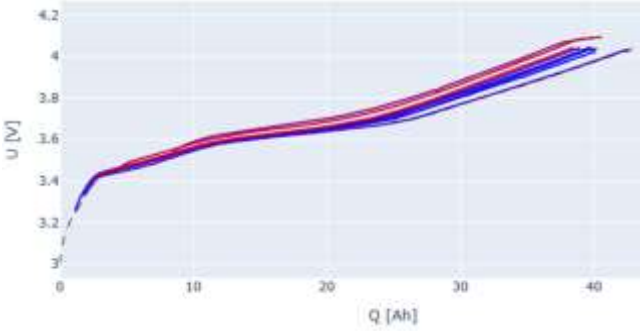


***Figure 2:** Measurement data of capacity checkup tests at different SOH. During ageing the distance to OCV curve (dashed line) is increasing.*

3 Estimation Algorithms

As outlined in the introduction section there are various elaborate ageing prediction models which are able to predict specific ageing processes. However, for application in field data analysis, empirical models are more suitable. Within this section, three methods with increasing complexity are presented. The first approach is based on public available data sheet parameter of the cell, the second one is based on laboratory ageing investigations. The third one also uses field data to enhance ageing prediction. In all methods the capacity is used as SOH indicator. This means other degradation modes like SEI growth or loss of active material are not considered for simplicity and clarity.

3.1 Baseline Estimation

A coarse SOH estimation can be performed using guaranteed mileage of the vehicle manufacturer. However, this value does not reveal information about the shape of the capacity fade during lifetime. This is illustrated in Figure 3. The shape of the curves can differ for different cell types as well as for different ageing modes. Thus, intermediate SOH prediction may be uncertain.
In addition, the ageing process only applies under specific conditions. These may differ significantly from the actual operating conditions.

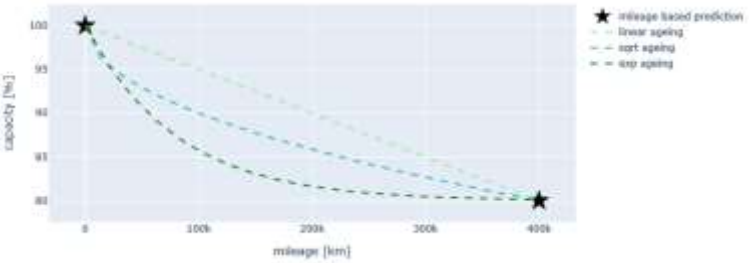


Figure 3: Illustration of various capacity curves over mileage.

3.2 Laboratory-Model based Estimation

Model based estimations not only consider charge throughput, but they also take into account the operation conditions. Typically, it is distinguished between calendric- and cyclic ageing. Calendric ageing describes the loss of capacity resulting from side reactions that take place also during idle phases. Cyclic ageing is caused by various degradation mechanism, like SEI growth or electrolyte decomposition. For calendric ageing only SOC and temperature are taken into account. For cyclic ageing typically also the current rate and the depth of discharge (DOD) is considered. Many models use the concept of weighting these input values. Hence the stress of a specific condition, like high temperature is quantified. This also applies to the model presented. It is based on the calculation of weighted full cycles w_q and weighted lifetime w_t:

$$w_q = \frac{1}{2C_{nom}} \int \alpha(SOC, I, T, DOD)\ |I|\ dt$$

$$w_t = \frac{1}{s_t} \int \beta(SOC, T) dt$$

The stressfactor functions α and β quantify the impact of the operation conditions. Accurate parametrization is a considerable challenge. Therefore, the methodology described in (14) is applied.
The two variables w_q and w_t are then used to calculate SOH. An ageing curve is used for this purpose, which describes the form of capacity loss. If no information for the cell under investigation is available, often a linear relationship is assumed:

$$c = c_0 - (c_0 - c_{EOL})\left(\frac{w_q}{l_q} + \frac{w_t}{l_t}\right)$$

l_q and l_t are the cycle life and calendar life parameter. c_0 denotes the initial capacity (typical 100%) and c_{EOL} the capacity at EOL (typical 80%).

Since chronological sequence of operation condition does not effect the SOH calculation, it is possible to use a histogram based approach. Thereby calendric and cyclic operation conditions are stored in multidimensional histograms. In Figure 4 the marginal 1-dimensional histograms for a bus of the analysed fleet are shown. It can be seen that the operation profile remains constant over the years. The temperature control keeps batteries in a comfortable region. The battery is mainly operating in the upper SOC region and the DOD is rather small.

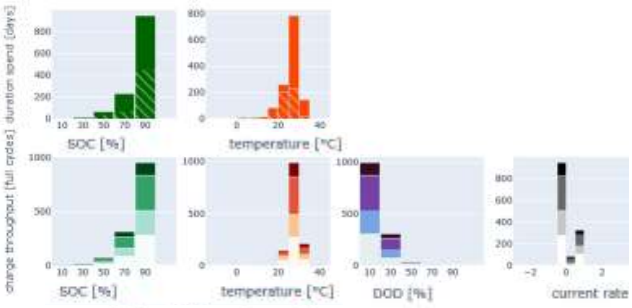


Figure 4: Histograms describing the operation of example battery.

In Figure 5 the resulting mean calendric- and cyclic stressfactors are shown. They are in the same region for almost all buses of the fleet. The calendric stress is approximately 1.0, indicating conditions similar to the reference case. The cyclic stress is even below 1, since the DOD is small. Thus, lifetime may exceed the expected value.

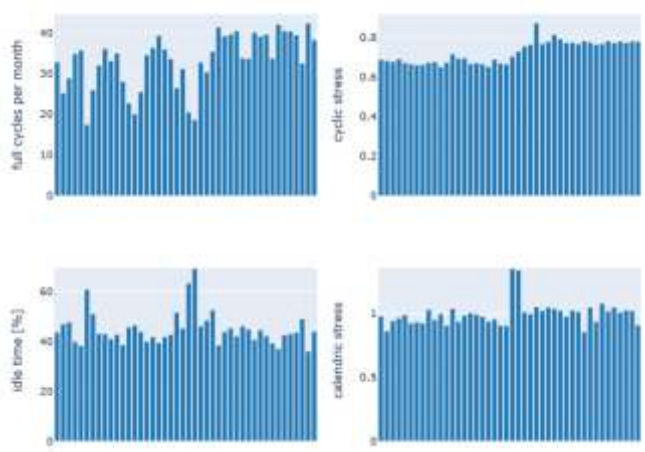


Figure 5: Monthly full cycles and idle time for all buses of the fleet as well as calendric and cyclic stress factors.

3.3 Field Data optimized Estimation

The approaches presented so far are based on laboratory tests of battery cells. However, these cannot reproduce all the effects that occur under real-world conditions. In (15), for example, it was shown that dynamic cycling can extend service life. In order to use field data to improve ageing models, several problems must be solved. To this end, the fleet model presented in (16) was developed. It enables ageing curves to be trained directly using measured voltage data. Additional time-consuming and costly reference tests are not necessary. The basic approach is shown in Figure 6.

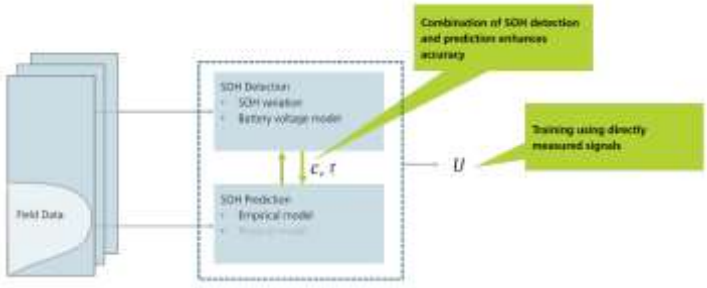


Figure 6: structure of the Fleet Model.

The fleet model combines ageing prediction with ageing detection, enabling joint learning from all fleet data. As a further advantage, in addition to the capacity loss, the increase in resistance can be determined.

4 Benchmark Results

The three presented approaches for predicting SOH were applied to the electric bus fleet. The results are evaluated using the dedicated capacity checkup tests.
The SOH estimation results of the fleet model (field data optimized estimation) are plotted in Figure 7. It appears that ageing does not proceed linearly. Instead, capacity drops sharply during the first full cycles. After that, the capacity fade rate slows down. At the end of the monitoring period the capacity of the batteries is in the range of 85%-90%. The resistance increased by only 15% at the

same time. The mean capacity prediction error is below 2% as can be seen in the boxplot.

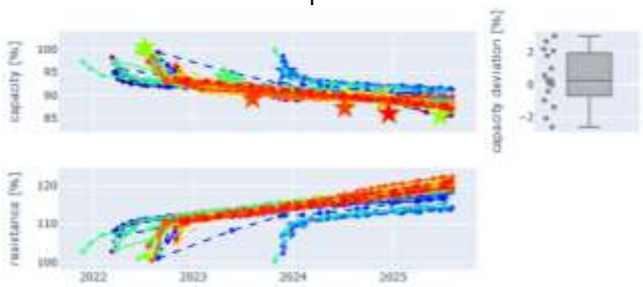


Figure 7: Capacity and resistance estimation using the field data enhanced estimation approach. Results of reference tests are plotted as star. The box plot on the right side shows the estimation error.

In Figure 8 also the results of the other methods are compared to the checkup tests. As expected, the accuracy of the baseline approach is the worst. However, even the model parameterized with laboratory data shows significant errors. Since it is not adapted to the specific cell, the capacity decrease is underestimated. This is mainly due to the nonlinear behaviour, i.e., the initial drop in capacity.

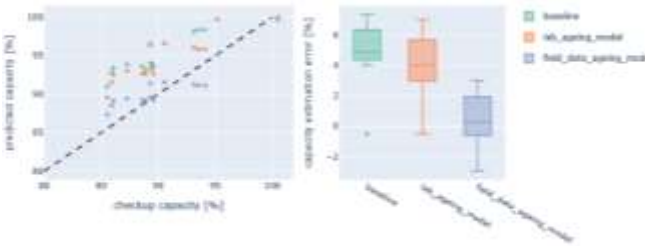


Figure 8: Benchmark of the analysed estimation methods. The left plot shows the estimated capacities compared to the checkup results. The box plot on the right shows the estimation error of each method.

5 Conclusions

The analysis of the electric bus fleet shows the importance of taking field data into account when determining ageing. Laboratory data alone cannot adequately reflect many effects under real-world conditions. Furthermore, the transfer between different cell types is questionable. These problems are solved by fitting the ageing model directly to the fleet data. Hence this benchmark study shows the importance of continuously monitoring and analysing battery data during its operation for reliable SOH prediction.

6 Literaturverzeichnis

1. *Degradation diagnostics for lithium ion cells.* **Christoph Birkl, Matthew R. Roberts, Euan McTurk, Peter G. Bruce, David A. Howey.** 2017, Journal of Power Sources, Bd. 341.
2. *Accelerated Aging Characterization of Lithium-ion Cells: Using Sensitivity Analysis to Identify the Stress Factors Relevant to Cyclic Aging.* **Lienkamp, Tanja Gewald and Adrian Candussio and Leo Wildfeuer and Dirk Lehmkuhl and Alexander Hahn and Markus.** 2020, MPDI Batteries, Bd. 6.
3. *Determination of degradation modes of lithium-ion batteries considering aging-induced changes in the half-cell open-circuit potential curve of silicon–graphite.* **Jossen, Julius Schmitt and Markus Schindler and Andreas Oberbauer and Andreas.** 2022, Journal of Power Sources, Bd. 532.
4. *Aging formula for lithium ion batteries with solid electrolyte interphase layer growth.* **Tanim, Tanvir R. und Rahn, Christopher D.** 2015, Journal of Power Sources, Bd. 294, S. 239-247.
5. *Perspective—Combining Physics and Machine Learning to Predict Battery Lifetime.* **Aykol, Muratahan und Storey, Chirranjeevi Balaji Gopal and Abraham Anapolsky and Patrick K. Herring and Bruis van Vlijmen and Marc D. Berliner and Martin Z. Bazant and Richard D. Braatz and William C. Chueh and Brian D.** 3, 2021, Journal of The Electrochemical Society, Bd. 168.
6. *Large-scale field data-based battery aging prediction driven by statistical features and machine learning.* **Qiushi Wang, Zhenpo Wang, Peng Liu, Lei Zhang, Dirk Uwe Sauer, Weihan Li.** 12, 2023, Cell Reports Physical Science, Bd. 4, S. 101720.
7. **Zhao, Nan Xu and Yu Xie and Qiao Liu and Fenglai Yue and Di.** A Data-Driven Approach to State of Health Estimation and Prediction for a Lithium-Ion Battery Pack of Electric Buses Based on Real-World Data. *MDPI: sensors.* 2022, S. 22.
8. **Marco, L.A. Román-Ramírez and J.** Design of experiments applied to lithium-ion batteries: A literature review. *Applied Energy.* 2022, S. 119305.
9. **Consortium, Battery Pass.** *Battery Passport Content Guidance.* thebatterypass.eu : thebatterypass.eu, 2023.
10. *Fleet Management Approach for Manufacturers displayed at the Use Case of Battery Electric Vehicles.* **Meisen, Friedrich von Bülow and Felix Heinrich and Tobias.** 2021, IEEE International Conference on Systems, Man, and Cybernetics.
11. *Electric Vehicle Batteries: Status and Perspectives of Data-Driven Diagnosis and Prognosis.* **Zhao, Jingyuan and Burke, Andrew F.** 8, 2022, Batteries, Bd. 10.
12. *The challenge and opportunity of battery lifetime prediction from field data.* **Howey, Valentin Sulzer.** 2021, Joule, Bd. 5.
13. *Empirical Energy Consumption Estimation and Battery Operation Analysis from Long-Term Monitoring of an Urban Electric Bus Fleet.* **Tom Klaproth, Erik Berendes,Thomas Lehmann,Richard Kratzing and Martin Ufert.** 8, 2025, World Electric Vehicle Journal, Bd. 16, S. 419.
14. *Learning the Ageing Behaviour of Lithium-Ion Batteries Using Electric Vehicle Fleet Analysis.* **Lehmann, Thomas and Berendes, Erik and Kratzing, Richard and Sethia, Gautam.** 12, 2024, Batteries, Bd. 10, S. 432.
15. *Dynamic cycling enhances battery lifetime.* **Onori, Alexis Geslin and Li Xu and Devi Ganapathi and Kevin Moy and William C. Chueh and Simona.** 2024, Nature Energy.

16. *Lithium-Ion Battery Aging Analysis of an Electric Vehicle Fleet Using a Tailored Neural Network Structure.* **Lehmann, Thomas and Weiß, Francis.** 7, 2023, Applied Sciences, Bd. 13.
17. **Zhao, Jingyuan and Burke, Andrew F.** Electric Vehicle Batteries: Status and Perspectives of Data-Driven Diagnosis and Prognosis. *Batteries.* 2022, S. 142.

DOI 10.5162/iCCC2026/1.3

Smart Sensor for continuous condition monitoring and predictive maintenance of a milling machine

Joris Liebermann, Martin Lehmann, Dirk Mayer, André Schneider, Akshaya Bindu Gowri, Fraunhofer IIS, Dresden, Germany
joris.liebermann@iis.fraunhofer.de

Abstract

We present a wireless, retrofittable smart sensor for continuous vibration-based condition monitoring and predictive maintenance of milling machines. The work is motivated by the need to detect specific conditions of the machine, like tool wear, to protect surface quality in high-throughput machining of large, costly parts. Our goal is to evaluate a low-cost hardware solution that can be quickly applied to existing milling machines against a state-of-the-art system. Following a methodical development cycle, we build a smart sensor including hardware and software components, acquire ground-truth data of milling with a sharp and a blunt tool, develop an edge-deployable AI model, implement the AI model on the smart sensor, and validate the implementation. Final deployment demonstrates reliable classification and actionable alerts, with comparative tests indicating monitoring quality comparable to a high-cost reference at reduced cost and integration complexity. Thus, the smart sensor can operate as a milling assistant, signalling worn tools with minimal installation effort and maximal data privacy.

1 Introduction

Damage or wear of milling tools causes deteriorated surface quality of produced components. When machining expensive parts or at high speed, early detection of these effects is mandatory to minimize rejects and optimize the process. This motivates in-process assessment of milling quality using machine-integrated sensing.

The application of sensors to detect tool wear has been extensively researched in the past. However, modern sensing technologies based on microelectronics and MEMS (micro-electromechanical systems) enable compact, wireless, and smart sensor systems, that can be realized at scale and deployed to a larger range of milling machines [1]. The implementation of advanced algorithms like neural networks on these devices even allow for complete offline processing, ensuring maximal data privacy and low latency. Generating information on tool wear requires the integration of advanced signal analyses and sensors into the machine. Among the most sensitive measurands are cutting force, vibration and acoustic emission or even a fusion of these [2], [3]. With the goal of developing a scalable solution, our study puts a focus on vibration sensing, which can be integrated with a moderate effort.

Wireless sensors alleviate retrofitting of monitoring systems to machines [4]. However, since vibration signal transmission requires high bandwidth, there is a motivation to integrate the signal analysis directly with the sensor and reduce transmission to relevant information needed for controlling the process.

Artificial intelligence (AI) has proven useful for tool wear analysis in the past, however has not been implemented in commercial products widely yet. A prerequisite for the transfer of such from lab to the shop floor is the automated operation of the monitoring system and a user interface that provides useful information to the machine operator [5]. Thus, our development targets the integration of AI based signal analysis directly at the sensor to provide tool condition information to the user.

With the introduction of Industry 4.0 and the industrial internet of things (IIOT), smart sensors, that integrate acquisition, processing and communication capabilities have gathered significant attendance [6]. Low-cost MEMS sensor hardware has been successfully qualified according to industrial standards [7]. Thus, consequently signal analysis should be implemented by low-cost systems, to enable scalable smart sensor systems for industrial use cases.

In the following sections of the paper, we introduce the applied methods, including smart sensor design, experimental set ups, AI model development and validation approach. After that, we present performance results obtained and close the report with a discussion and conclusion.

2 Methods

2.1 Design Methodology

For building the smart vibration sensor we followed a previously introduced development cycle [8], that was applied before [9] allowing for efficient iterative design and rapid prototyping. The five steps of this methodical approach are as follows: 1. Ground truth data acquisition from the machine, 2. Data analysis of the obtained data, 3. Engineering of the smart sensor, 4. Validation in lab tests, 5. Integration into the target environment. This workflow allows all subsequent development, testing and optimization after a single data acquisition campaign.

2.2 System Architecture

The hardware base of the smart sensor is an Arduino Nano 33 BLE Sense Rev2 as it is a compact cost-effective microcontroller board with BLE capabilities and acceleration sensor. The board is equipped with a nRF52840 (Arm® Cortex®-M4F, 256kB RAM) possessing sufficient computational power and memory to run fully quantized neural

This work was partly funded by the Free State of Saxony within the Center for Prototyping and Test of AI systems.

networks. The on-board BMI270 is a digital MEMS-based Inertial Measurement Unit (IMU) and provides three axis acceleration measurements with up to 1600 samples/s and a selectable range of +-2g to +-16g.
For wireless operation, we integrated a Li-Ion battery with charging module. A 3D printed enclosure shields the electronic system against environmental impacts like cooling liquid and manual handling. Two magnets on the base enable for rapid mounting on the target surface of a machine. The board is screw-mounted to the bottom of the housing with a tight mechanical coupling to the bottom magnets, ensuring optimal transmission of the vibration signals from the machine into the sensor element. As the system should also work offline, an external LED was added for immediate feedback.
A custom sensor configuration for the BMI270 was uploaded to achieve the maximal sampling rate, by outputting raw data instead of processed g-values to speed up reading operation. The First-In-First-Out buffer of the IMU ensured continuous monitoring while processing or sending data. The maximal data output rate of 1600 samples/s was set at a range of +-4g.
The obtained raw data in three axes is batched into maximum sized Bluetooth Low Energy payloads and send out with the minimal connection interval to a Raspberry Pi 4 that structures and publishes it to a server.

2.3 Experimental Setup and Data Acquisition

To evaluate the smart sensor under real operation conditions, we conducted face milling of a flat steel workpiece. The objective was to capture vibration data from different tool conditions, sharp and blunt, with the smart sensor and a reference system and compare their performance.
The milling program executed two opposing roughing passes. First, the sharp tool machined from the right edge toward the center and machined the surface line after line. After an automatic tool change, the blunt tool mirrored this process from the left edge towards the center. To isolate the passes from each other, an uncut bridge in the center is left in between the milled surfaces. Cooling liquid was applied throughout the whole process. The total runtime was approximately 4 min, obtaining 1min 46s of active milling data for each tool condition. The parameters of the milling process are summarized in Table 1.

Parameter	Value
Process	Conventional face milling
Tool	10 mm end mill (6 cutting edges)
Radial depth of cut	60 %
Axial depth of cut	2 mm
Spindle speed	3500 rpm
Feed rate	600 mm/min

Table 1: Parameters of Milling Process

As reference for the sensor performance a state-of-the-art system, the SKALI.KIT data acquisition system (developed at Fraunhofer IIS and Fraunhofer IZFP) was used to acquire vibration data from piezoelectric sensors and store it in the server. The accelerometer considered for comparison was a uniaxial Kistler KS77C.100 at a sampling rate of 16kHz.
Sensors were magnetically attached to a shoulder near the machine spindle. For evaluating the smart sensor performance, the reference data was aligned during postprocessing by down-sampling with a low-pass filter, DC removal and temporal alignment through linear interpolation. The comparison of the sensors occurred in both time and frequency domain.

2.4 AI Model & Embedded Algorithms Development

The vibration time series data acquired from the smart sensor was analyzed to identify meaningful features to use as input for the neural network. The most common features for vibration data in time domain and frequency domain were inspected. Frequency-domain analysis was prioritized over time-domain features due to clearer wear signatures and embedded deployment feasibility.
To mitigate mounting misalignment, only the z-axis spectrum (normal to mounting plane) was processed, which is not affected by in-plane rotations of the sensor.
For a robust classification, we employ a compact frequency-domain feature pipeline that is also suited for smart sensor deployment. From 512 samples, we apply Hanning windows with 50% overlap and compute the magnitude of the Fourier spectrum at a size of 64 bins as computing power and memory are limited in the Arduino. Given the previously obtained time series we yield a data set with 663 training samples for each label.
A fully connected neural network (TensorFlow Keras) consisting of 64 input neurons, two hidden layers and a single output neuron was trained with 75:25 train/test split and Adam optimizer, using binary cross entropy for 40 epochs. Deployment of the neural network in the smart sensor required conversion to TensorFlow Lite for quantization to 8-bit. The resulting model was translated to a C header file and implemented into the firmware of the Arduino along with the same feature pipeline. In the final smart sensor, the classification results were signaled by the external LED that activated at detection of a worn tool.

2.5 Validation in the Laboratory

For testing the proper implementation of the signal processing pipeline in the microcontroller with respect to the Python implementation that was used during model development, a “known answer test” is conducted. After training and testing the algorithms on the PC, a prerecorded time series was processed both on the microcontroller and on the PC and the outputs of the feature generation pipeline were compared. Large differences in the feature vectors would indicate implementation errors. Also, the output of the AI model using the same features as input is compared.
To evaluate the hardware implementation, the recorded vibration time series obtained during milling were replayed on an electrodynamic shaker with 20kHz bandwidth and a closed control loop. Due to the stationary character of the

signals and the low latency of the implementation, 6 seconds for each of the signals are sufficient to test the performance of the proposed smart sensor.
For the final deployment at the machine, the smart sensor with classification firmware was placed in the same location as in the data acquisition phase and the milling program was restarted. As accuracy metric for reliable classification the LED was tracked visually during the whole process.

3 Results

3.1 Measurements

Figure 1 displays the spectrogram of the acceleration data in z axis obtained by the Arduino. During operation of both tools the signal exhibits significant excitation at the same harmonically spaced frequency bands, while there are low-energy segments in-between caused by transition of the toolpath. The sharp tool causes lower magnitudes with excitation at very defined frequency bands, while the blunt one exhibits higher excitation with less defined transitions. The active milling process is visibly divided by the reduced signal magnitude during tool change.

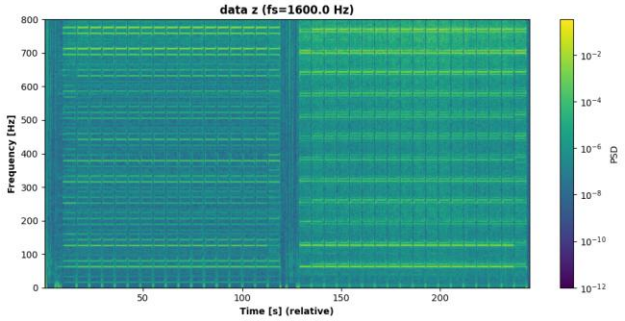


Figure 1: Spectrogram of acceleration in z direction acquired by the smart sensor displaying milling with a sharp tool (left) and blunt tool (right)

3.2 Data Comparison

Figure 2 presents the synchronized acceleration data from the smart sensor and the reference sensor (KS77C.100) over the first 5 seconds. Both signals show considerable agreement in the time domain, with similarities at transient events. After beginning of milling operation at about 3s, the amplitude of the smart sensor is slightly higher than the reference sensor.

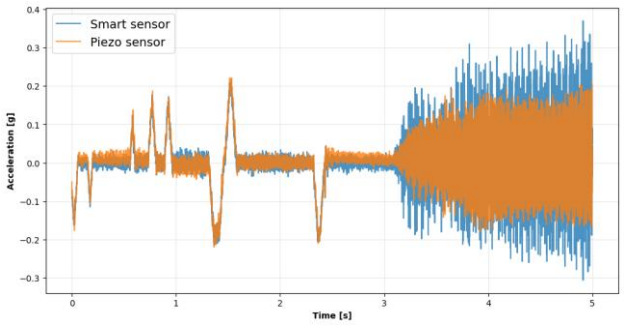


Figure 2: Acceleration data obtained from the smart sensor (blue) and the piezo sensor (orange)

The differences in power spectral density of the smart sensor to the reference sensor during the milling operation are depicted in Figure 3. Both capture the strong excitation at the specific frequency band mentioned before, while the Arduino exhibits a larger magnitude baseline and less variation across the frequency spectrum.

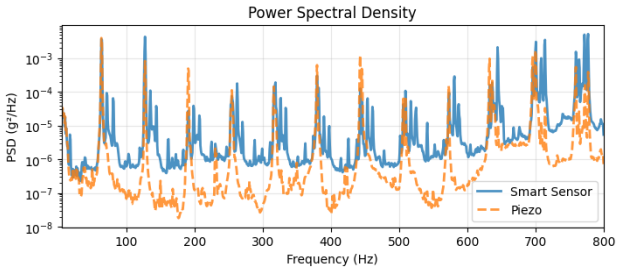


Figure 3: Power spectral density of piezo (orange) and smart sensor (blue) during milling process

3.3 AI model Performance

After training on the frequency spectrum data set, the neural network achieved a classification accuracy of 97% on the test set of 166 samples for each tool condition. Quantization of the model to 8-bit integers did not reduce the classification accuracy on the test set and produced identical results as before.

	True sharp	**True blunt**
Predicted sharp	156	0
Predicted blunt	10	166

Table 2: Confusion Matrix

Accordingly, the confusion matrix in Table 2 demonstrates that all blunt tool samples were detected correctly while a small amount of sharp tool data was classified as blunt. The only outliers are 10 false positives, where sharp tool vibrations are misclassified as blunt during low-energy segments when the toolpath changed during the active milling process.

3.4 Validation

The comparison of calculated feature vectors in Python to the microcontroller implementation only exhibited minor differences with a mean relative deviation of 0.006% between with a maximum difference of 0.022%. The resulting classification matched. During the shaker tests, the smart sensor produced correct classifications in real-time for replay of both the blunt and the sharp tool data segments. In the final deployment at the machine, the smart sensor continuously achieved reliable classification through the whole milling process. Only two incorrect classifications were observed during the sharp tool usage, whereas all blunt-tool operations were detected correctly.

4 Discussion

The results demonstrate that our five-stop development methodology is effective for rapid prototyping of a smart vibration sensor for tool condition monitoring. The system

reliably distinguishes between sharp and worn tools in both lab tests and real-world applications.
The model has a small bias of occasionally predicting a sharp tool as blunt during short low-energy transition segments when there is no active cutting, as this occurs for both tool conditions. Validation shows that the feature pipeline is correctly implemented, with only minor deviations probably due to differences in numerical precision of the systems.
Through rigorous validation on our shaker test bed, performance is comparable to traditional sensor setups with data processing on a dedicated acquisition and processing unit or in the cloud.
Evidentially, low-cost hardware is generally able to capture the differences in vibration signature accurately enough to distinguish between the different tool conditions. Although the raw measurement accuracy is slightly below that of industrial grade piezo sensors, the entire system performance outweighs their advantages and allows for flexible adaptation and configurability. While traditional sensor setups often require complex cable management and bulky data acquisition hardware, our battery-driven solution exhibits quick time-to-measurement with minimal setup effort. The embedded algorithms and AI models potentially eliminate the need for any further data processing depending on the use case. However, the use of a battery also entails higher maintenance effort since the battery needs to be charged from time to time even at the most conservative energy saving modes. For a long-term monitoring setup, a wired setup with constant energy supply should be considered. Nevertheless, in contrast to traditional sensor setups, energy could be supplied to multiple sensors with a single (daisy-chain) cable reducing wiring to a minimum.
Limitations of the presented solution contain limited data variance as we only conducted a few experiments. Therefore, the results are limited to exactly this setup. No generalization over milling with other materials, cutting parameters or machines can be applied. This will be the target of ongoing research in this field.

5 Conclusion

This work successfully developed and validated a wireless, retrofittable smart sensor system for vibration-based tool condition monitoring in milling machines. The Arduino-based solution, featuring edge-deployed AI processing, achieved 97% classification accuracy in distinguishing between sharp and worn tools while operating entirely on-device. Careful and rigorous design and validation demonstrated performance comparable to high-cost commercial reference systems despite using low-cost MEMS sensors. Key advantages include minimal installation effort through magnetic mounting, complete data privacy due to on-sensor processing, and reduced integration complexity compared to traditional wired monitoring systems. The validation process confirmed reliable real-world performance with only two false classifications during extended operation. While the current implementation is specific to the tested machining parameters and would require retraining for different setups, this approach represents a practical, cost-effective solution for predictive maintenance that significantly lowers the barrier for implementing condition monitoring on existing milling equipment. Future work should address expanding the model's generalizability across different materials and machining parameters while maintaining the system's simplicity and ease of deployment.

6 Literature

[1] Teti, R. ; Mourtzis, D. ; D'Addona, D.M. ; Caggiano, A.: Process monitoring of machining. In: *CIRP Annals* Bd. 71 (2022), Nr. 2, S. 529–552
[2] Mohanraj, T. ; Shankar, S. ; Rajasekar, R. ; Sakthivel, N.R. ; Pramanik, A.: Tool condition monitoring techniques in milling process — a review. In: *Journal of Materials Research and Technology* Bd. 9 (2020), Nr. 1, S. 1032–1042
[3] Mey, Oliver ; Schneider, André ; Enge-Rosenblatt, Olaf ; Mayer, Dirk ; Schmidt, Christian ; Klein, Samuel ; Herrmann, Hans-Georg: Condition Monitoring of Drive Trains by Data Fusion of Acoustic Emission and Vibration Sensors. In: *Processes* Bd. 9 (2021), Nr. 7, S. 1108
[4] Zhang, Cunji ; Yao, Xifan ; Zhang, Jianming ; Jin, Hong: Tool Condition Monitoring and Remaining Useful Life Prognostic Based on a Wireless Sensor in Dry Milling Operations. In: *Sensors* Bd. 16 (2016), Nr. 6, S. 795
[5] Bombiński, Sebastian ; Kossakowska-Skajster, Joanna ; Jemielniak, Krzysztof: New developments and future prospects in commercial tool condition monitoring systems. In: *Measurement* Bd. 255 (2025), S. 118037
[6] Schütze, Andreas ; Helwig, Nikolai ; Schneider, Tizian: Sensors 4.0 – smart sensors and measurement technology enable Industry 4.0. In: *Journal of Sensors and Sensor Systems* Bd. 7 (2018), Nr. 1, S. 359–371
[7] Iqbal, Ali ; Mian, Naeem S. ; Longstaff, Andrew ; Fletcher, Simon: Performance Evaluation of Low-Cost Vibration Sensors in Industrial IoT Applications. In: *International Journal of Automation Technology* Bd. 16, Fuji Technology Press Ltd. (2022), Nr. 3, S. 329–339
[8] Mayer, Dirk ; Lehmann, Martin ; Beyer, Volkhard ; Schneider, Andre ; Schneider, Peter: Challenges in Design and Validation of MEMS Based Smart Sensor Systems. In: *2024 Symposium on Design, Test, Integration and Packaging of MEMS/MOEMS (DTIP)*, 2024, S. 1–6
[9] Lehmann, Martin ; Liebermann, Joris ; Mayer, Dirk ; Beyer, Volkhard: Performance And Energy Consumption Of Smart Sensors For Vibration-Based Anomaly Detection. In: *2025 Smart Systems Integration Conference and Exhibition (SSI)*, 2025, S. 1–6

DOI 10.5162/iCCC2026/1.4

Berstdruckvorhersage von faserverstärkten Druckbehältern anhand von Schallemissionen

Burst pressure prediction with Acoustic Emissions on composite pressure vessels

Emanuel D. Kästle, Ali Ghaznavi, Eric Duffner
Bundesanstalt für Materialforschung und -prüfung (BAM), 3.5 Sicherheit von Gasspeichern und Gefahrguttanks, Berlin, Deutschland
emanuel-david.kaestle@bam.de

Kurzfassung

Neue Arbeiten zeige eine Methode zur Vorhersage der Versagenslast von kohlenstofffaserverstärkten Polymerwerkstoffen (CFRP) mittels Schallemission auf. Dies ermöglicht eine zerstörungsfreie Bewertung der Restfestigkeit von Druckbehältern und trägt so zur Verbesserung der Betriebssicherheit bei. Der hier betrachtete Datensatz umfasst sieben kleinformatige CFRP-Proben unter Zugbelastung sowie fünf vollständige CFRP-Druckbehälter vom Typ IV. Alle Prüfkörper wurden in periodischen Lade- und Entladezyklen getestet, um charakteristisch Kennwerte bei unterschiedlichen Laststufen zu erfassen. Auf Basis dieser Daten wird ein neuronales Netz trainiert, das die Kennwerte mit den Laststufen relativ zur Versagenslast verknüpft. Die Ergebnisse zeigen, dass Vorhersagen auch über unterschiedliche Prüfkörper hinweg möglich sind; Modelle, die mit Probendaten trainiert wurden, können den Berstdruck vollständiger Behälter vorhersagen – und umgekehrt. Einschränkungen im Hinblick auf potenzielle Verzerrungen und die Notwendigkeit aussagekräftiger Unsicherheitsabschätzungen werden diskutiert und bleiben eine Herausforderung für zukünftige Arbeiten.

Abstract

Recently, a method for predicting the failure load of carbon fiber reinforced polymer (CFRP) materials has been presented. The method relies on passive acoustic emission (AE) monitoring during loading. This approach could be used in the non-destructive evaluation of the residual strength of pressure vessels and therefore improve the operational safety. A test dataset is obtained from 7 small-scale CFRP specimens, subjected to tensile load, as well as 5 full CFRP type IV pressure vessels. All test objects are equipped with AE sensors and periodically loaded and unloaded to get characteristic AE-related ratios at different load steps. Using these data, we train a neural network to link AE-derived ratios to load levels relative to the failure load. It is shown that predictions are possible, also across different test objects, i.e. trained with specimen data, burst pressure predictions for full vessels are feasible and vice versa. We discuss limitations related to potential biases and the need for meaningful uncertainty estimation, which remain open for future work.

1 Introduction

The global energy system is transitioning toward sustainability. In this context, hydrogen has gained prominence as a versatile energy carrier, offering a pathway to decarbonize sectors that are difficult to electrify, such as heavy industry, long-haul transportation, and seasonal energy storage [1]. Compressed hydrogen (up to 700 bar) is widely used for onboard storage in fuel cell vehicles and transport trailers. To achieve safe and efficient storage under these conditions, type IV carbon fiber reinforced polymer (CFRP) pressure vessels have become the industry standard [2]. These composite vessels offer an exceptional strength-to-weight ratios, corrosion resistance, and durability under cyclic loading. Since pressure vessels represent a safety critical element of the hydrogen value chain, they must undergo regular inspections or must be operated with a pre-defined limited-service lifetime. Inspections are costly and often require destructive testing of sample vessels. Vessels that have reached the end of their service lifespan, may still fulfill necessary safety criteria so that disposal implies a waste of valuable resources. A major factor for the costs and CO2 emissions of CFRP vessels is the energy-intense carbon fiber production [2]. At the same time, recycling options are still very scarce or inefficient. The BAM is therefore interested in developing a method for non-destructive testing of CFRP vessels to improve the operational safety and to avoid long out-of-service periods during inspection. Recently, Sause et al. [3, 4] have presented a method for the failure load prediction of CFRP specimen and shown that it is also applicable on full vessel scale. This method is based on testing with acoustic emissions (AE), a non-destructive testing method based on the occurrence of ultrasonic signals when a material is subjected to a force. At BAM, we have conducted preliminary tests on several CFRP specimens under tensile load and on two different sizes of type IV CFRP vessels from different manufacturers. The results of these test will be presented, and the accuracy of the failure load prediction will be discussed. The terms failure load (for specimens) and burst pressure (for vessels) are used interchangeably throughout this work.

DOI 10.5162/iCCC2026/1.4

2 Method

The applied method for the load failure prediction is based on the works of Sause et al. [3, 4]. The following summarizes the method with minor deviations. AE data are recorded during alternating loading/unloading to compute Felicity [5], Shelby, and energy-related ratios (Fig. 1).

Following Sause et al. [3], the Felicity ratio is calculated from a percentage threshold of the total number of hits recorded during one load/pressure step. A hit describes one short ultrasonic burst signal, emitted by the material and recorded at the AE sensors. The Felicity ratio is therefore given by:

$$FR_{n\%} = L_{n\%}/L_0$$

The variable n gives the threshold value so that $L_{5\%}$ is the load level where the threshold of 5% of the signals is exceeded. Different ratios at thresholds of 5, 10, 15 and 20% are extracted from the data [3]. A fifth ratio is given by the mean of the four aforementioned ones, FR_{mean}. The Shelby ratio (SR) is based on a similar concept as the FR, but counts the number of hits during unloading, with threshold levels at 80, 85, 90 and 95% of the total number of hits. A fifth ratio will be SR_{mean}. The third type of AE-based ratios are the energy (ENE), signal strength (SST) and amplitude (AMP) ratios. The definition of these parameters follows the one implemented in the Vallen software (www.vallen.de). The energy ratio is given as the load level where the cumulated energy of the current load increase, E_1, exceeds 80% of the cumulated energy of the previous load step, E_0, relative to the previous load maximum, L_0:

$$ER = L_{E_1 \geq 0.8E_0}/L_0$$

The 0.8 factor, absent in [3], prevents undefined ratios when current energy does not exceed previous energy. The amplitude and signal strength ratios are calculated analogously. This results in a total of 13 ratios which are used in the burst pressure prediction approach.

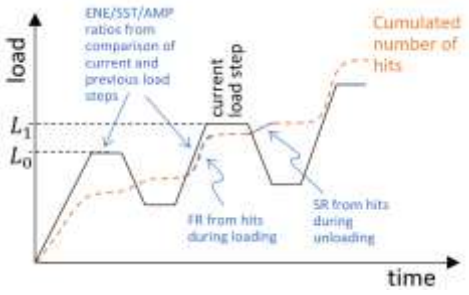


Figure 1 Schematic loading and unloading profile used in this work. FR and SR are based on the cumulated number of hits during loading and unloading. The other ratios are based on cumulated energies, signal strength and amplitudes.

As shown in Fig. 1, there is no complete unloading after each load maximum, which is different from [3]. The load is decreased by a factor of 0.5. This means that the smallest possible ratios are 0.5. The largest possible value for the Shelby ratios is always 1, which follows from the definition above. For the Felicity and ENE/SST/AMP ratios, it is given by L_{i+1}/L_i. In our tests, $L_{i+1} - L_i$ is kept constant over the load profile, so that the largest possible value for the ratio decreases with increasing load.

Ratios are input to a two-layer neural network [3]. The network has 13 inputs (or 8 without Shelby ratios), one hidden layer (10 nodes), and an output layer with two nodes, representing the mean and the standard deviation of the load ratio (LR) which is defined as:

$$LR = L_i/L_{failure}$$

where L_i is the current load level (Fig. 1). Using the LR instead of absolute values makes the model more generalizable. The LR values required for model training are given from the true failure loads (burst pressures) of the tested specimens (vessels). A single prediction could estimate failure load, but this introduces high uncertainty. Therefore, several predictions should be made, from which the average failure load is calculated by a linear regression (Fig. 2). The slope is computed via weighted linear regression using LR standard deviations as weights. The lower and upper boundaries of the failure load are estimated from the mean slope ±2 standard deviations.

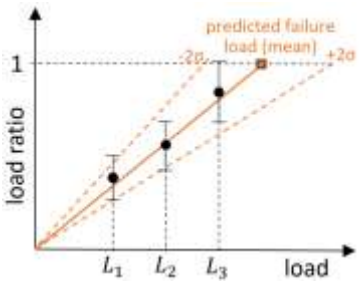


Figure 2 Schematic failure load prediction from three predicted LRs . The dashed orange lines indicate slope lines ±2 times standard deviations (σ).

3 Data

The dataset includes seven CFRP specimens with alternating 0°/90° fiber layers, tested under tensile load. The free length of samples (between the reinforcement tabs) is 15 cm, the width is 2.5 cm and the thickness approx. 2 mm (Fig. 3). Each sample is monitored with four Vallen VS900-M broadband sensors connected to Vallen AEP5 preamplifiers and a Vallen AMSY-6 data logger to record the AE signals during testing. Additionally, five vessels are tested, each with 15 identical sensors and the same measurement setup and AE equipment. Three sensors are glued onto each dome and nine sensors are glued with approximately equal distance onto the cylindrical surface of the vessels using hot melt adhesive. Two small vessels (~50 cm × 15.5 cm, labeled ‘S’) and three large vessels (~90 cm × 23.5 cm, labeled ‘L’) were hydraulically tested.

An example of the extracted ratio data is given in Fig. 4. The Felicity and energetic ratios clearly follow a trend, which is usable for predictions, while the extracted Shelby ratios are more scattered. The variability may be due to the lower number of hits during unloading phases. Therefore, the Shelby ratios will not be used in the following. Also, Sause et al. [4] did not use Shelby ratios in their more recent work. On the one hand, the FR and ER trends depend

DOI 10.5162/iCCC2026/1.4

on the load levels, since the largest possible values are L_{i+1}/L_i. On the other hand, the trends are controlled by a change in AE activity at different load levels. The dataset comprises ratio data at load ratios close to 1, i.e. the failure load. This is, because tests on specimens and on vessels S1 and S2 were conducted until failure with acoustic emission sensors attached. For the L1 to L3 vessels, the tests were stopped at a pressure close to bursting and the sensor were removed. The burst test was conducted only afterwards. The obtained failure loads for the CFRP specimens are shown in Fig. 4.

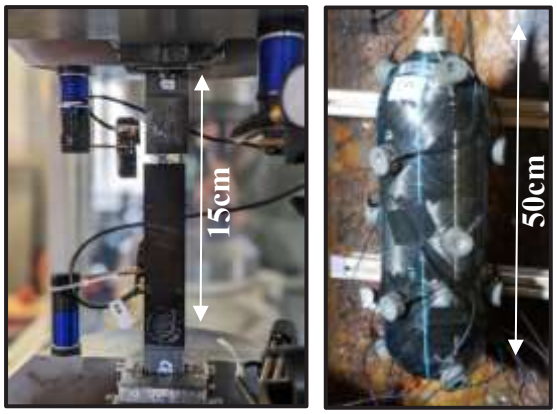


Figure 3 Test setup for tensile specimen testing (left) and for a full vessel of type S (right). 15 sensors are glued to the surface of the vessel.

The shown ratio values are calculated from all available sensors taken together. This gives more stable values and reduces the effect from outliers. However, it is also possible to treat the data for each sensor separately which can add to the estimations of the prediction uncertainty or can be used to identify regions with higher AE activity [4]. Despite the factor of 0.8 for the ENE/SST/AMP ratios, on some occasions, no valid values can be extracted where E_l<E_0 (*ER* equation above). A valid datapoint includes all 5 FRs and 3 energetic ratios. The total number of valid datapoints is 88.

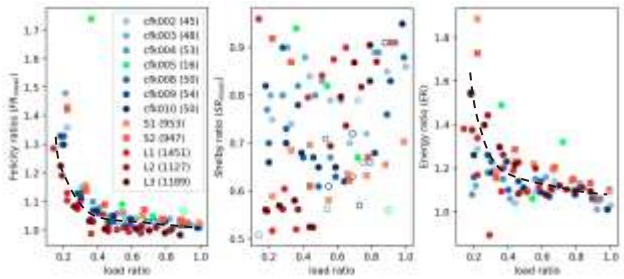


Figure 4 FR_{mean}, SR_{mean} and ER data extracted from the load/pressurization tests. The other extracted ratios are not shown for the sake of clarity. Open circles/squares indicate incomplete datapoints, because of missing ER values. Labels *cfk* correspond to tensile specimen data, *L* and *S* describe large and small vessel data. The average data trends are highlighted with dashed lines. The failure loads (in kN) and the burst pressures (in bar) are given in brackets in the legend behind each sample. The sample data highlighted in green is discussed in section 5.

4 Results

Several results are presented, for different combinations of training and testing datasets. Since the number of datapoints is relatively small, the test dataset also serves as validation dataset. Training is stopped when the validation residual starts being consistently above the training residual. With larger datasets, a separation of validation and test datasets, and a validation dataset taken from the training data, should be preferred.

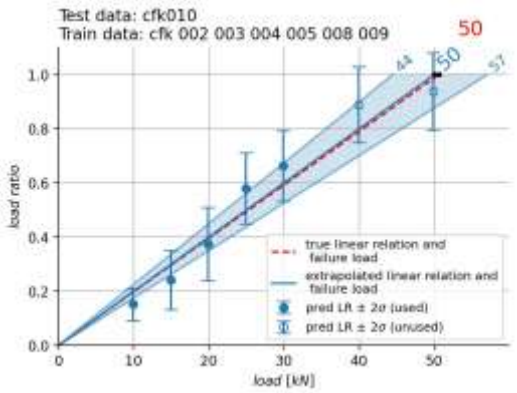


Figure 5 Failure load prediction for sample cfk010 with a model trained on the remaining 6 cfk samples.

Figure 5 gives an example only using CFRP specimen data. For this example, the prediction (49 kN) is very close to the true failure load at 50 kN. In the diagram, all load ratio predicted by the neural network are plotted together with their estimated standard deviation. In the following linear regression step, the last two datapoints are not used, since in an actual application, the test would be stopped, and a prediction would be made before critical loads are reached. In this and all the following examples, we do not consider predicted load ratios above 70% of the true failure load in the linear regression. We choose this threshold to simulate a realistic test which is stopped at a safe margin before failure.

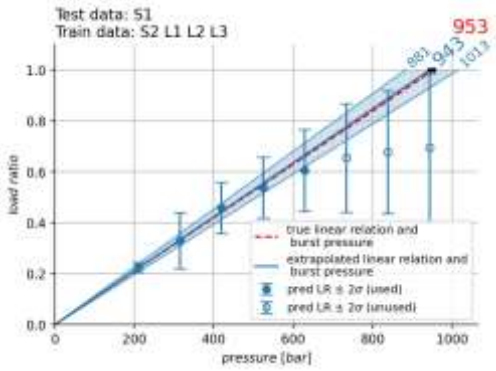


Figure 6 Failure load (burst pressure) prediction for vessel S1 with a model trained on the remaining 4 vessels.

In Figure 6, a similar test is presented using only vessel data. The predicted burst pressure is about 55 bar below the true burst pressure. An offset in the LRs is visible, more pronounced at larger pressures. This could lead to

DOI 10.5162/iCCC2026/1.4

larger errors if more LR values (beyond the 70% limit) were used in the linear regression. This is likely to be explained with the offset in the S- and L-vessel data, where ratios extracted for the S vessels consistently show slightly higher ratios (Fig. 4). A reason for this is discussed below.

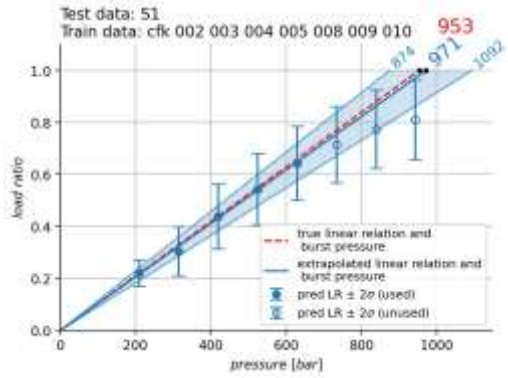


Figure 7 Failure load (burst pressure) prediction for vessel S1 with a model trained on CFRP tensile specimen.

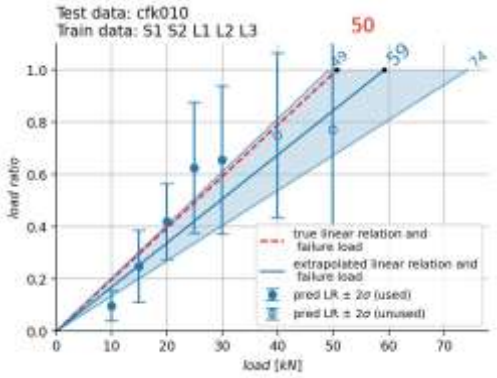


Figure 8 Failure load prediction for sample cfk010 with a model trained on 5 vessel datasets.

Finally, Figs. 7 and 8 present mixed prediction results for training datasets based on specimen data and predictions for vessel data, and vice versa. The predicted failure loads / burst pressures are within the given uncertainty margins in both cases. This shows that predictions are possible across the different test objects, even though the uncertainties may be high.

5 Discussion

The results presented in the preceding section show that the failure load prediction method works and can even be applied across different CFRP sample types. Given the similar behavior of the ratio data presented in Fig. 4, a positive result was to be expected. There are, however, some limitations that deserve attention. Currently, the dataset is relatively small, so that the results would change if only a few additional data points were introduced. This means that the predictions, including the predicted uncertainties are not very reliable, yet. A proper dataset should probably include serval dozens of tested specimens, ideally subjected to other tests (e.g., bending) and with different layups and defects. A much larger number and variety of vessels would further help to estimate the reliability of the method. With the current approach of averaging data from all sensors, local weak zones may go undetected (e.g., from impacts), so that obtaining ratios separately for different sensors, or even localized to certain regions on the vessel surface could be beneficial [4].

The uncertainties, estimated from the slope ±2σ only give a very basic guess for the true uncertainty. The resulting uncertainties only depend on how well the point align on a linear relation, without taking the predicted load ratio standard deviations into account. A more rigorous approach could be provided by a Bayesian linear regression.

The load/pressure release is 50% in every step (Fig. 1), which is different from Sause et al. [3] who unload complete. This results in longer testing times, but also more signals, and larger spread in the Felicity, Shelby and energetic ratios from low to high loads. The fact that the Felicity ratios hardly ever go below 1 in this work is related to the limited unloading in combination with the pristine condition of the specimens/vessels that have never been loaded before: Most signals are emitted when high load levels are reached for the first time. The larger spread from complete unloading could therefore help to improve the prediction accuracy. Small deviations from the planned loading profile can significantly influence the obtained ratios. This includes shorter/longer holding phases between loading and unloading and deviations towards too high or too low loads which did happen in our experiments, e.g. by small leakages or overshooting of the pressurization device. A very severe influence is further observed if loading is repeated, e.g. a vessel which has already seen pressures up to 1000 bar, and then ratio data is extracted for different steps below that level. This results in ratios that deviate strongly from the profile shown in Fig. 4. The current method is therefore only applicable for pristine CFRP materials.

For the CFRP vessel datasets, there appears to be a small offset in the ratio data between S and L vessels (Fig. 4). It is also evident that a similar offset occurs for the one sample with significantly lower failure load (cfk005). This can be easily understood when considering that the x-axis in Fig. 4 gives the load ratio. At a given load step, a reduced failure load results in a higher load ratio so that all datapoints are shifted to the right. As shown in Fig. 6, this can cause a bias in the predicted values. This phenomenon is thus related to the fact that the observed trend in the ratio data is mainly caused by having fixed $L_{i+1} - L_i$ steps and consequently a reduction in the L_{i+1}/L_i ratios with higher load. A correction for this effect or additional inputs to the neural network might therefore be desirable to obtain more reliable results.

6 Conclusions

An implementation of the failure load prediction method first introduced by Sause et al. [3] is presented. By comparing data from CFRP specimens under tensile load, as well as from two types of CFRP pressure vessels, it is

shown that good predictions are possible, even across different test object types and small sample sizes. The presented dataset is relatively small (7 specimens, 5 vessels), nevertheless the true failure loads are always within a ±2σ range from the predicted values. Future developments should take more diverse CFRP tests and specimens into account, improve the estimation the prediction uncertainty, and consider a careful re-evaluation of the input ratios which are, in parts, too scattered (Shelby ratios) or can produce prediction biases (Felicity, energetic ratios).

7 Acknowledgements

We want to thank the project partners and colleagues involved in the DAVID project and at BAM who conducted many of the required tests. This work was supported by the German Federal Ministry of Economic Affairs and Energy within the DAVID project (03LB3099C) and the BTU-BAM Graduate School „Trustworthy Hydrogen“.

8 References

[1] S. O. Akpasi, I. M. S. Anekwe, E. K. Tetteh, U. O. Amune, S. I. Mustapha und S. L. Kiambi, „Hydrogen as a clean energy carrier: advancements, challenges, and its role in a sustainable energy future,“ *Clean Energy,* Bd. 9, pp. 52-88, January 2025.

[2] T. Q. Hua, R. K. Ahluwalia, J.-K. Peng, M. Kromer, S. Lasher, K. McKenney, K. Law und J. Sinha, „Technical assessment of compressed hydrogen storage tank systems for automotive applications,“ *International Journal of Hydrogen Energy,* Bd. 36, pp. 3037-3049, February 2011.

[3] M. Sause, S. Schmitt und S. Kalafat, „Failure load prediction for fiber-reinforced composites based on acoustic emission,“ *Composites Science and Technology,* Bd. 164, p. 24–33, August 2018.

[4] M. G. R. Sause, S. Schmitt, B. Hoeck und A. Monden, „Acoustic emission based prediction of local stress exposure,“ *Composites Science and Technology,* Bd. 173, pp. 90-98, March 2019.

[5] T. J. Fowler, „Acoustic Emission of Fiber Reinforced Plastics,“ *Journal of the Technical Councils of ASCE,* Bd. 105, pp. 281-289, 1979.

DOI 10.5162/iCCC2026/2.2

Integrative Bewertung von Mikrozirkulationsstörungen bei Post-Covid 19 mittels Hyperspektralbildgebung und kardiopulmonaler Leistungsdiagnostik

Autoren

Lea Schlierkamp-Voosen[1], Barlo Hillen[1,2], Vincent Weber[1], Aleksandar Tomaskovic[1], David Tobias Ochmann[1], Severin Zentgraf[1], Daniel Andrés López[3], Elmo W.I. Neuberger[1], Perikles Simon[1], Eva-Verena Griemert[4]

Affiliation

[1]Institut für Sportwissenschaft, Abteilung für Sportmedizin, Prävention und Rehabilitation, Johannes Gutenberg-Universität, Mainz, Deutschland
[2]Institut für Arbeits-, Sozial- und Umweltmedizin, Universitätsmedizin der Johannes Gutenberg-Universität, Mainz, Deutschland
[3]Institut für Informatik, Forschungsgruppe Computational Geometry, Johannes Gutenberg-Universität, Mainz, Deutschland
[4] Klinik für Anästhesiologie, Universitätsklinikum, Johannes Gutenberg-Universität, Mainz, Deutschland

Kontakt: lschlier@students.uni-mainz.de

Kurzfassung

Bei Patienten mit Post-Covid-19 Syndrom (PCC) zeigen sich Hinweise auf relevante Mikrozirkulationsstörungen, die mittels Hyperspektralbildgebung objektivierbar sind. In der Studie unterschieden sich PCC-Betroffene bereits in Ruhe statistisch signifikant von Kontrollpersonen hinsichtlich der Gewebesauerstoffsättigung (StO_2) und des Nahinfrarot-Perfusions-Index (NPI), insbesondere an den Handflächen (StO_2: p=0,017; NPI: p=0,042). Unter Belastung zeigten sich deutliche gruppenspezifische Veränderungen, vor allem bei den Pre-Post-Differenzen von StO_2 und NPI an den Unterarmen (StO_2: p=0,004; NPI: p=0,007). Zudem korrelierten diese mikrovaskulären Anpassungsreaktionen positiv mit der kardiorespiratorischen Fitness (VO_2peak (StO_2: ρ=0,575, p=0,012 ; NPI: ρ=0,633, p=0,005)), VO_2@VT1 (StO_2: ρ=0,707, p=0,001 ; NPI: ρ=0,497, p=0,036) und VO_2@VT1/VO_2predicted (StO_2: ρ=0,546, p=0,019), was auf einen Zusammenhang zwischen belastungsinduzierter Mikrozirkulationsveränderung und körperlicher Leistungsfähigkeit hindeutet. Die Ergebnisse bestätigen das Potenzial der Hyperspektralbildgebung in Kombination mit der Spiroergometrie als diagnostisches Instrument zur Detektion und Quantifizierung von Mikrozirkulationsstörungen bei PCC.

1 Hintergrund/Einleitung

Das Post-Covid-19 Syndrom (PCC) beschreibt Beschwerden, die länger als 12 Wochen nach akuter SARS-CoV-2-Infektion persistieren. Die Symptomatik ist vielfältig und umfasst ein breites Spektrum physischer, psychischer und kognitiver Beeinträchtigungen. Besonders häufig sind Fatigue, Dyspnoe, Muskel- und Gelenkschmerzen, posttraumatische Belastungsstörungen, Angst- und Schlafstörungen, Depressionen sowie kognitive Dysfunktionen [1]. Als zentrale Ursache werden unter anderem Störungen der Mikrozirkulation, einschließlich endothelialer Dysfunktion, Kapillarverlust und der Bildung von Mikrothromben beschrieben. Dies führt zu einer verminderten Gewebeoxygenierung. Dringender Bedarf besteht an objektiven diagnostischen Verfahren zur Erfassung jener pathophysiologischen Mechanismen [2,3,4]. Ziel dieser Arbeit ist die explorative Untersuchung der Hyperspektralbildgebung im Rahmen der Spiroergometrie zur Quantifizierung von Veränderungen der Mikrozirkulation bei PCC.

2 Methoden

Es wurden 18 PCC-Patienten untersucht und mit gematchten Kontrollpersonen verglichen. Zur Gewinnung der Hyperspektralbilder nutzten wir das TIVITA Tissue Kamera-System (Diaspective Vision GmbH, Am Salzhaff, Germany), wie in Abbildung 1 dargestellt. Das Prinzip der hyperspektralen Bildgebung beruht auf den spezifischen Eigenschaften von oxygeniertem und desoxygeniertem Hämoglobin sowie von Wasser bezüglich Absorption und Reflexion von Licht. Die TIVITA-Kamera verwendet hierbei Wellenlängen zwischen 500 und 1000 Nanometern zur nichtinvasiven Erfassung der vier Parameter StO_2 (Gewebesauerstoffsättigung [%]), NPI (Nahinfrarot-Perfusions-Index [Arbitrary Units]), THI (Gewebehämoglobin-Index [Arbitrary Units]) und TWI (Gewebewasser-Index [ArbitraryUnits]), welche die lokale Sauerstoffversorgung des Gewebes abbilden. Die visuelle Aufbereitung der Daten sowie die quantitative Auswertung über manuell definierte,

kreisrunde Regions of Interest (ROI) erfolgte mittels der kameraeigenen Software TIVITA Tissue Suite, welche farbkodierte Bilder der gemessenen Parameter generiert (siehe Abbildungen 2 und 3). Die Aufnahmen wurden an beiden Handflächen und Unterarmen der Probanden jeweils vor und nach einer standardisierten Belastungsuntersuchung auf dem Fahrradergometer mit Atemgasanalyse durchgeführt. Im Rahmen dieser Spiroergometrie wurden unter anderem die Parameter VO_2peak [ml/min/kg], VO_2@VT1 [ml/min/kg], VO_2@VT1/VO_2predicted [%], VO_2peak/VO_2predicted [%] und Peak Power Output [W/kg] erhoben, welche der objektiven Beurteilung der Leistungsfähigkeit der Probanden dienen. Zur Validierung des Verfahrens der Hyperspektralbildgebung wurde zudem bei 26 PCC-Betroffenen eine kapilläre Blutgasanalyse (BGA) vor und nach der Belastung durchgeführt.

Die statistische Auswertung wurde mit IBM SPSS Statistics durchgeführt. Zur Prüfung der Voraussetzungen für parametrische Testverfahren wurde zunächst die Verteilung der Daten mithilfe des Shapiro–Wilk-Tests bewertet. Abhängig von der Verteilungsform und der Art der Stichprobe kamen unterschiedliche Testverfahren zum Einsatz. Bei normalverteilten Daten wurden für Mittelwertsvergleiche innerhalb der Gruppen (PCC-Patienten bzw. Kontrollprobanden) mit gepaarten t-Tests und Gruppenvergleiche mit unabhängigen t-Tests durchgeführt. Für nicht-normalverteilte Daten wurden entsprechende nichtparametrische Verfahren angewendet, der Wilcoxon-Signed-Rank-Test für abhängige Stichproben sowie der Mann–Whitney-U-Test für unabhängige Stichproben. Zur Untersuchung von Zusammenhängen zwischen hyperspektralen Parametern, leistungsphysiologischen Variablen sowie Parametern der Blutgasanalyse wurde der Spearman-Rangkorrelationskoeffizient (ρ) berechnet. Die Stärke der Korrelationen wurde gemäß etablierten Klassifikationen interpretiert, wobei Werte von 0,00–0,10 als vernachlässigbar, 0,10–0,39 als schwach, 0,40–0,69 als moderat, 0,70–0,89 als stark und 0,90–1,00 als sehr stark interpretiert wurden [5]. Ein zweiseitiges Signifikanzniveau von $p < 0,05$ wurde für alle Analysen als statistisch signifikant definiert.

3 Ergebnisse

Die Messungen vor der Belastung zeigten bei den PCC-Betroffenen für den Parameter StO_2 der Handfläche einen Mittelwert (MW) von 58,9% (± 7,6%) und des Unterarms von 44,9% (± 4,0 %). Der Parameter NPI wies an denselben Messpunkten MW von 55,5 (± 6,0) und 44,3 (± 5,1) auf. Die MW des THI betrugen 39,8 (± 8,3) und 24,7 (± 7,0), während TWI MW von 61,3 (± 2,2) und 45,3 (± 1,8) zeigte. In der Kontrollgruppe lagen die MW des StO_2 vor Belastung an der Handfläche bei 65,8% (± 9,1%) und am Unterarm bei 45,2% (± 2,7%). Die NPI-MW betrugen 59,6 (± 5,7) und 44,4 (± 2,7). Der Parameter THI wies MW von 41,8 (± 6,6) und 22,2 (± 6,6) auf, während die MW des TWI an denselben Messpunkten 62,2 (± 2,2) und 47,6 (± 2,2) betrugen. Die MW der Messparameter für Hand und Unterarm beider Gruppen im Ruhezustand sind in Tabelle 1 und 2 zusammengefasst.

Parameter	PCC-Betroffene	Kontrollprobanden	p-Wert
StO_2	58,9% (± 7,6%)	65,8% (± 9,1%)	0,017
NPI	55,5 (± 6,0)	59,6 (± 5,7)	0,042
THI	39,8 (± 8,3)	41,8 (± 6,6)	0,436
TWI	61,3 (± 2,2)	62,2 (± 2,2)	0,243

Tabelle 1 TIVITA-Parameter Pre-Werte (Handfläche links)

Parameter	PCC-Betroffene	Kontrollprobanden	p-Wert
StO_2	44,9% (± 4,0%)	45,2% (± 2,7%)	0,584
NPI	44,3 (± 5,1)	44,4 (± 2,7)	0,839
THI	24,7 (± 7,0)	22,2 (± 6,6)	0,267
TWI	45,3 (± 1,8)	47,6 (± 2,2)	0,002

Tabelle 2 TIVITA-Parameter Pre-Werte (Unterarm links)

Im Vergleich der Werte vor und nach Belastung zeigte sich bei den PCC-Betroffenen an der Handfläche eine signifikante Zunahme der Parameter StO_2 ($p<0,001$) und NPI ($p=0,010$) (siehe Tabelle 3). Ebenso ließen sich für die zwei Parameter auch in der Kontrollgruppe signifikante Unterschiede zwischen den Zeitpunkten vor und nach Belastung feststellen, an der Handfläche (StO_2: $p=0,049$; NPI: $p=0,043$) sowie am Unterarm (StO_2: $p<0,001$; NPI: $p=0,024$), wie in den Tabellen 4 und 6 dargestellt.

Parameter	Pre	Post	p-Wert
StO_2	58,9% (± 7,6%)	64,6% (± 6,7%)	<0,001
NPI	55,5 (± 6,0)	58,4 (± 4,6)	0,010
THI	39,8 (± 8,3)	38,4 (± 7,7)	0,435
TWI	61,3 (± 2,2)	61,3 (± 2,4)	0,927

Tabelle 3 TIVITA-Parameter Pre-Post-Vergleich: PCC-Betroffene (Handfläche links)

Parameter	Pre	Post	p-Wert
StO_2	66,2% (± 9,8%)	71,6% (± 4,7%)	0,049
NPI	59,9 (± 5,9)	63,9 (± 4,1)	0,043
THI	39,9 (± 5,7)	40,5 (± 6,5)	0,660
TWI	62,6 (± 2,0)	62,2 (± 2,1)	0,270

Tabelle 4 TIVITA-Parameter Pre-Post-Vergleich: Kontrollprobanden (Handfläche rechts)

Parameter	Pre	Post	p-Wert
StO_2	44,9% (± 4,0%)	45,5% (± 7,0%)	0,417
NPI	44,3 (± 5,1)	44,5 (± 5,7)	0,693
THI	24,7 (± 7,0)	23,6 (± 8,3)	0,311
TWI	45,3 (± 1,8)	45,1 (± 2,3)	0,432

Tabelle 5 TIVITA-Parameter Pre-Post-Vergleich: PCC-Betroffene (Unterarm links)

Parameter	Pre	Post	p-Wert
StO_2	45,2% (± 2,7%)	50,9% (± 4,8%)	<0,001
NPI	44,4 (± 2,7)	46,4 (± 3,4)	0,024
THI	22,2 (± 6,6)	21,0 (± 7,2)	0,480
TWI	47,6 (± 2,2)	47,6 (± 2,3)	1,000

Tabelle 6 TIVITA-Parameter Pre-Post-Vergleich: Kontrollprobanden (Unterarm links)

Im Vergleich der beiden Gruppen waren für StO_2 und NPI im Ruhezustand signifikante Divergenzen an den Handflächen zu beobachten (StO_2: $p=0,017$; NPI: $p=0,042$) (siehe Tabelle 1). Zudem ergaben sich für die Pre-Post-Differenzen dieser Parameter an den Unterarmen hochsignifikante Unterschiede zwischen den Gruppen (StO_2: $p=0,001$; NPI: $p=0,032$), wie in Tabelle 8 ersichtlich.

Parameter	PCC-Betroffene	Kontrollprobanden	p-Wert
StO_2	5,76% (± 6,01%)	3,85% (± 10,72%)	0,308
NPI	2,87 (± 4,18)	2,24 (± 6,40)	0,542
THI	-1,43 (± 5,41)	0,17 (± 6,84)	0,444
TWI	-0,03 (± 1,69)	-0,19 (± 1,52)	0,784

Tabelle 7 TIVITA-Parameter Pre-Post-Differenzen (Handfläche links)

Parameter	PCC-Betroffene	Kontrollprobanden	p-Wert
StO_2	0,67% (± 3,38%)	5,80% (± 4,95%)	0,001
NPI	0,30 (± 2,14)	2,17 (± 2,83)	0,032
THI	-1,30 (± 4,45)	-1,70 (± 6,96)	0,836
TWI	-0,17 (± 1,12)	-0,22 (± 1,71)	0,909

Tabelle 8 TIVITA-Parameter Pre-Post-Differenzen (Unterarm links)

Im Rahmen der Spiroergometrie zeigten sich bei den PCC-Betroffenen für die Parameter VO_2@VT1 (MW = 8,6 ml/min/kg (± 2,1); Cut-Off < 11 ml/min/kg) und VO_2@VT1/VO_2predicted (MW = 38,0 % (± 8,9); Cut-Off < 45%) Werte deutlich unter den Normbereichen. Es stellte sich ein moderater positiver signifikanter Zusammenhang zwischen den Pre-Werten von StO_2 und NPI an der Handfläche und dem Parameter VO_2@VT1/VO_2predicted heraus (StO_2: ρ=0,559, p=0,016; NPI: ρ=0,670, p=0,002). Darüber hinaus korrelierten die Pre-Post-Differenzen beider Parameter am Unterarm deutlich mit VO_2peak (StO_2: ρ=0,575, p=0,012; NPI: ρ=0,633, p=0,005) sowie mit VO_2@VT1 (StO_2: ρ=0,707, p=0,001; NPI: ρ=0,497, p=0,036). Der Zusammenhang zwischen der Differenz von StO_2 und VO_2@VT1 wird in Abbildung 4 dargestellt. Für VO_2@VT1/VO_2predicted ließ sich ebenfalls ein positiver signifikanter Zusammenhang mit den Differenzen von StO_2 am Unterarm beobachten (ρ=0,546, p=0,019).

Die MW der Blutgasanalyse-Parameter pH, pO_2, pCO_2 und sO_2 lagen in den Normbereichen. Lediglich der Laktatwert wies nach Belastung eine deutliche Erhöhung (MW = 4,12 mmol/l (± 2,45); Normbereich: 0,6-1,7 mmol/l) auf. Zwischen den Ruhewerten der BGA-Parameter und den spiroergometrischen Leistungsparametern bestanden keine signifikanten Zusammenhänge. Es zeigte sich jedoch ein moderater positiver signifikanter Zusammenhang zwischen den Pre-Post-Differenzen von pO_2 und sO_2 und dem Parameter VO_2@VT1/VO_2predicted (pO_2: ρ=0,445, p=0,026; sO_2: ρ=0,540, p=0,005). Zudem korrelierten VO_2peak und VO_2peak/VO_2predicted positiv mit der Pre-Post-Differenz des Laktatwerts (VO_2peak: ρ=0,596, p=0,001; VO_2peak/ VO_2predicted: ρ=0,649, p=<0,001) sowie negativ mit der Differenz des pH-Werts (VO_2peak: ρ=-0,524, p=0,006; VO_2peak/ VO_2predicted: ρ=-0,409, p=0,038).

4 Schlussfolgerung/Diskussion

Die Hyperspektralbildgebung erweist sich als vielversprechendes Verfahren zur objektiven Erfassung mikrovaskulärer Veränderungen bei PCC-Betroffenen. Als besonders aussagekräftige Parameter konnten StO_2 und NPI identifiziert werden, insbesondere aufgrund der Robustheit gegenüber Positions- und Seitenunterschieden und der deutlichen belastungsinduzierten Veränderung. Die signifikanten Gruppenunterschiede zwischen PCC-Betroffenen und Kontrollpersonen sowie die positiven Korrelationen der belastungsinduzierten Veränderungen von StO_2 und NPI mit den spiroergometrischen Leistungsparametern deuten zudem auf einen Zusammenhang zwischen mikrovaskulärer Anpassung und körperlicher Leistungsfähigkeit hin. Die ergänzend durchgeführte kapilläre Blutgasanalyse bestätigt diese Befunde, indem sie belastungsabhängige Veränderungen von pO_2, sO_2, Laktat und pH zeigt, die mit zentralen Leistungsparametern signifikant korrelieren. Insgesamt legen die Ergebnisse nahe, dass die Hyperspektralbildgebung in Kombination mit der Spiroergometrie ein wertvolles diagnostisches Instrument zur objektiven Charakterisierung des PCC und einhergehender Mikrozirkulationsstörungen darstellen kann.

DOI 10.5162/iCCC2026/2.2

5 Abbildungen

Abbildung 1 TIVITA Tissue Hyperspektralkamera

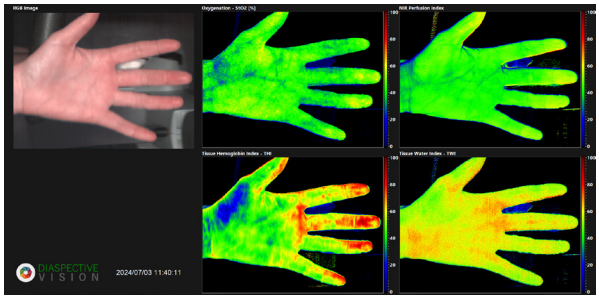


Abbildung 2 Visuelle Aufbereitung der Daten mittels TIVITA Tissue Suite-Software

Abbildung 3 Quantitative Auswertung der Gewebesauerstoff–sättigung (StO_2) mit drei ROIs

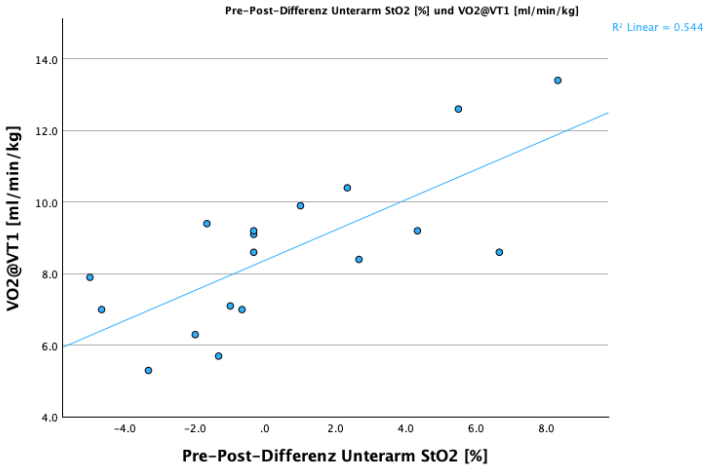


Abbildung 4 Scatter Plot: Pre-Post-Differenz von StO_2 [%] und VO_2@VT1 [ml/min/kg]

6 Literatur

[1] Luo D, Mei B, Wang P, Li X, Chen X, Wei G, et al. Prevalence and risk factors for persistent symptoms after COVID-19: a systematic review and meta-analysis. Clin Microbiol Infect. 2024;30(3):328-35.

[2] Koutsiaris AG. A blood supply pathophysiological microcirculatory mechanism for Long COVID. Life (Basel). 2024;14(9).

[3] Schäfer H, Teschler M, Mooren FC, Schmitz B. Altered tissue oxygenation in patients with post COVID-19 syndrome. Microvasc Res. 2023;148:104551.

[4] Kuchler T, Günthner R, Ribeiro A, Hausinger R, Streese L, Wöhnl A, et al. Persistent endothelial dysfunction in post-COVID-19 syndrome and its associations with symptom severity and chronic inflammation. Angiogenesis. 2023;26(4):547-63.

[5] Schober P, Boer C, Schwarte LA. Correlation coefficients: appropriate use and interpretation. Anesth Analg. 2018;126(5):1763-1768

DOI 10.5162/iCCC2026/2.3

Mobility Monitoring for Geriatrics: Gait Detection and Parameter Extraction using an e-Textile Garment IMU and Deep Learning.

1st Tobias Steinmetzer
Department of Therapy Sciences (II)
Brandenburg University of Technology Cottbus-Senftenberg
Cottbus, Germany
tobias.steinmetzer@b-tu.de

2nd Sven Michel
Department of Therapy Sciences (II)
Brandenburg University of Technology Cottbus-Senftenberg
Cottbus, Germany

***Abstract*—This study introduces and validates a practical, two-stage pipeline for the unobtrusive detection and analysis of gait using a single Inertial Measurement Unit (IMU) integrated into a smart undershirt, positioned at the abdomen, aimed at enabling reliable long-term geriatric monitoring. Data were collected from 15 healthy young adults performing free movement and quiet standing. The pipeline first employs a 1D Convolutional Recurrent Neural Network (1D-CRNN) classifier using six IMU signals and derived features to segment gait versus standing activity. Subsequently, a routine utilizing bandpass filtering and peak detection on the Euler angle magnitude is used to count steps and derive cadence and step variability. The 1D-CRNN achieved excellent classification performance with an Accuracy of 0.981 $\pm$ 0.022 and a PR_AUC of 0.995 $\pm$ 0.005. The step detection successfully isolated continuous walking segments, yielding plausible cadence estimates (e.g., 94.6 steps/min) and measures of step variability (0.236 s). The smart-textile-based system provides a robust and lightweight solution for reliable gait detection and the extraction of fundamental mobility metrics from an unobtrusive trunk location. This foundation is crucial for developing practical wearable technologies for assessing functional health in older adults, with future work focusing on clinical validation in geriatric cohorts and expanding the activity classification spectrum.**

***Index Terms*—Gait Analysis, Smart Textiles, Inertial Measurement Unit (IMU), Convolutional Recurrent Neural Network (CRNN), Activity Classification, Mobility Monitoring**

Introduction

Gait is one of the most informative indicators of functional mobility and overall health in older adults [1]. Characteristics such as cadence, variability, and step symmetry are vital for assessing frailty, fall risk, and cognitive decline in geriatric populations [2], [3]. In clinical practice, slow or unstable gait is often an early predictor of reduced independence, risk of hospitalization, or increased mortality in elderly individuals [1].

Shin et al. (2025) demonstrated significant associations between sensor-derived gait metrics (such as cadence and step length) and clinical measures of muscle mass and physical performance in community-dwelling seniors [4].

From a methodological perspective, Oudre et al. developed a robust algorithm for automatic gait event detection (heel strikes, toe-offs) in both healthy subjects and severely impaired patients using filtered IMU acceleration and pattern matching [5]. In a similar vein, Stenger et al. (2025) validated an approach for gait event detection and step-length estimation using a single waist-mounted IMU, achieving high accuracy compared to a GAITRite reference system [6].

Together, these studies underscore both the clinical relevance and technical feasibility of using wearable IMUs for real-world gait monitoring in older adults, motivating our work that combines activity classification with step detection in a non-invasive, smart-textile-based setup. To address these gaps, this work introduces a novel, two-stage pipeline for unobtrusive gait analysis. Our main contributions are: 1) The development and validation of a 1D-CRNN model for accurate classification of gait versus standing using data from a single abdomen-mounted IMU integrated into a smart undershirt. 2) The reliable extraction of clinically relevant gait parameters (cadence, step count, and variability) from the segmented gait episodes, suitable for future geriatric mobility assessment.

Materials and Methods

Data set

The study protocol for the development of the smart undershirt was submitted to and favorably assessed by the Ethics Committee of the Brandenburg University of Technology Cottbus-Senftenberg on August 30, 2024 (File Number: EK2024-26).

The investigation included healthy adult participants. Inclusion criteria required participants to be able to ambulate without assistance. As the test was entirely voluntary, the primary exclusion criterion was participant unwillingness to proceed.

The current dataset, which represents a partial evaluation of the larger four-part experimental design, comprises recordings from a total of 15 participants (4 males and 11 females). The participants had a mean age of $22.27 years$ ($\sigma = 3.31 years$).

The data presented here were collected based on two phases:

1) Free-Movement Phase: Participants were first instructed to engage in unrestricted movement within a defined

DOI 10.5162/iCCC2026/2.3

Fig. 1. Prototype of the smart undershirt.

space for a period of **five minutes (5 minutes)**. Participants performed a variety of movements commonly encountered in Activities of Daily Living (ADL), such as walking backward, ascending/descending stairs, and normal ambulation. Participants were granted full creative freedom, provided the movements remained representative of routine daily activities.

2) Quiet Standing Phase: Subsequently, participants were instructed to stand still for a duration of five minutes (5 minutes).

The smart undershirt was designed to capture comprehensive physiological and movement data. It was equipped with a diverse array of integrated sensors [8]. For this analysis we use the Bosch/CEVA BNO085 IMU Sensor on Abdomen [7]. We recorded Euler angle and linear Acceleration with a frequency at 80 Hz, see Figure 1.

Activity Classifier

We convert quaternions to Euler angles (ZYX: yaw, pitch, roll) using standard rotation conventions and SciPy's spatial transform routines. Zero quaternions (all zeros) are replaced with the identity quaternion to avoid numerical errors. A time axis is generated with a frequency of 80 Hz. Channels used as model input: $yaw, pitch, roll, a_x, a_y, a_z$. For each sliding window we compute derived features: first differences of Euler and acceleration, acceleration magnitude, and Euler magnitude. The final per-sample feature vector contains the original 6 signals plus the first differences (6), plus two norms, resulting in 14 features per time step.

We segment each labeled region using a sliding window of length W samples and step size S. Each window inherits the label of the region it falls into (gait or standing). To avoid data leakage between subjects/recordings we perform group-wise cross-validation using GroupKFold, where groups correspond to file identifiers.

The RCNN architecture consists of two Conv1D layers (32 and 64 filters, kernel size 5) with max-pooling, followed by an LSTM (64 units), dropout (0.5), a dense layer (32 units) and a single-unit sigmoid output for binary classification. Binary cross-entropy loss and class weighting (computed from training folds) handle label imbalance.

We train with Adam optimizer and early stopping (patience can be added). Evaluation metrics include accuracy, precision, recall, F1-score, ROC-AUC and average precision (PR-AUC). GroupKFold (K=5) provides unbiased estimates across recordings.

Step Detection

For windows classified as gait, the step detection pipeline is:

1) Compute Euler angle magnitude: $e_{mag} = \sqrt{yaw^2 + pitch^2 + roll^2}$.
2) Bandpass filter e_{mag} within $0.5-3Hz$ using a 4th-order Butterworth filter and zero-phase filtering (filtfilt).
3) Smooth the filtered signal with a moving average with size 5 ($0.0625s$) to reduce residual noise.
4) Detect peaks with *scipy.signal.find_peaks* using adaptive criteria: minimal distance (e.g., 0.3 f_s samples), prominence proportional to the signal standard deviation, and optional height threshold.

This procedure produces peak timestamps corresponding to detected steps. From these we compute step count, inter-step intervals, cadence (steps per minute), and step-time variability (std of intervals).

RESULTS

Activity Classification

Table I summarizes the classification performance of the proposed 1D-CRNN model for detecting gait versus standing across different window sizes. All evaluated window configurations demonstrate high performance, with accuracy consistently above 0.97. Among them, the window size of 150 samples ($\approx$ 1.875 s) yielded the best overall performance across all metrics. Specifically, it achieved the highest F1 score, precision, and recall, indicating a strong balance between sensitivity and specificity.

Notably, the ROC_AUC and PR_AUC scores for the 150-sample window outperform the smaller window sizes by a significant margin, highlighting the model's superior ability to distinguish between gait and standing activities, even under class imbalance. These results suggest that slightly longer temporal windows allow the model to better capture the dynamic characteristics of gait patterns while maintaining generalization stability, as reflected by the low standard deviations across folds.

DOI 10.5162/iCCC2026/2.3

TABLE I
CLASSIFICATION PERFORMANCE OF THE 1D-CRNN MODEL WITH DIFFERENT WINDOW SIZES FOR 5 FOLDS. VALUES ARE REPORTED AS MEAN ± STANDARD DEVIATION. THE BEST PERFORMANCE WAS ACHIEVED FOR A WINDOW SIZE OF 150 SAMPLES ($\approx 1.875s$) (HIGHLIGHTED IN BOLD).

Metric	Window Size 100	Window Size 110	Window Size 150
Accuracy	0.979 ± 0.023	0.981 ± 0.023	**0.981 ± 0.022**
Precision	0.967 ± 0.045	0.968 ± 0.043	**0.970 ± 0.042**
Recall	0.994 ± 0.003	0.997 ± 0.003	**0.995 ± 0.003**
F1 Score	0.980 ± 0.022	0.981 ± 0.022	**0.982 ± 0.021**
ROC_AUC	0.985 ± 0.021	0.985 ± 0.021	**0.995 ± 0.004**
PR_AUC	0.979 ± 0.029	0.980 ± 0.028	**0.995 ± 0.005**

Step Detection and Stride Estimation

The proposed model successfully segmented the gait activity into three distinct walking episodes, as shown in Table II and Figure 2. Segment 1 represents the longest continuous gait interval with a duration of 145.88 s, during which 230 steps were detected, corresponding to a cadence of 94.6 steps/min. The step variability in this segment was 0.236 s, indicating a relatively stable walking pattern with moderate temporal consistency.

Segment 2 covers a shorter walking interval of 69.75 s and includes 113 steps, resulting in a cadence of 97.2 steps/min. The step variability (0.229 s) is comparable to Segment 1, suggesting similar gait regularity despite the reduced duration.

Segment 3 is a very short gait episode lasting only 1.88 s, containing 4 detected steps with a cadence of 128.0 steps/min. This unusually high step frequency and low variability (0.120 s) likely indicate a brief acceleration or transitional movement rather than steady-state gait.

Overall, the results confirm that the model reliably identifies gait segments of varying durations and accurately estimates key step-related metrics such as cadence and step variability. Longer segments demonstrated more stable gait patterns, whereas very short gait intervals exhibited higher variability and were less representative of typical walking behavior.

TABLE II
SEGMENT-WISE GAIT DETECTION AND STEP CHARACTERISTICS.

Metric/ Segments	1	2	3
Start (s)	5.94	153.69	468.69
End (s)	151.81	223.44	470.56
Duration (s)	145.88	69.75	1.88
Steps	230	113	4
Steps/min	94.6	97.2	128.0
Step Variability (s)	0.236	0.229	0.120

DISCUSSION

This study successfully validates a two-stage pipeline for unobtrusive mobility assessment using a single abdomen-mounted IMU integrated into a smart undershirt. The results confirm the robustness of the activity classification and the plausibility of the extracted gait metrics.

The proposed 1D-CRNN model demonstrated excellent performance in distinguishing gait from standing, achieving an accuracy above 0.98 and a PR_AUC of 0.995 (for W=150 samples, Table I). This performance is highly competitive and validates the use of the abdominal sensor placement in a textile format, which is key for high user acceptance in long-term geriatric monitoring [1]. The superior performance of the 150-sample window suggests that this duration effectively captures the dynamic characteristics of trunk movement during ambulation. The low standard deviations across cross-validation folds indicate strong model generalization within the healthy cohort.

The subsequent step detection pipeline, based on filtering the Euler angle magnitude, reliably extracted clinically relevant parameters like cadence and step variability (Table II). The mean cadence (94.6 to 97.2 steps/min) is consistent with typical walking speeds for healthy adults, and the low variability (≈0.23 s) confirms a stable, steady-state gait. The results demonstrate the pipeline's capability to provide basic mobility parameters critical for assessing frailty and fall risk [3].

The clear segmentation of activity into "Gait" and "Standing" is not only valuable for mobility assessment but also provides distinct periods for extracting additional physiological data. Notably, the identified quiescent "Standing" phases can be leveraged for highly accurate measurements of other vital signs. This is reinforced by our parallel work, which has demonstrated the successful application of a 1D-CRNN for respiration monitoring using the same smart textile platform during periods of low activity [9]. This integration capability underscores the potential of the smart undershirt as a truly multi-functional monitoring system.

While effective, the current study has limitations: The data was collected exclusively from young adults, requiring validation on the target geriatric population where gait patterns and sensor signals are expected to differ significantly [2]. Furthermore, the extracted step metrics require validation against a gold standard system to confirm their absolute clinical accuracy, and the activity scope must be extended to a broader range of Activities of Daily Living (ADL).

CONCLUSION

We presented a practical and reproducible two-stage pipeline for gait detection and step analysis utilizing data from a single abdomen-mounted Inertial Measurement Unit (IMU) integrated into a smart undershirt. The 1D-CRNN model achieved an excellent activity classification accuracy (Accuracy > 0.98) between gait and standing, demonstrating the robust capability of the textile-integrated sensor setup. Following classification, key gait parameters such as cadence and step variability were reliably extracted, making the approach suitable for assessing fundamental mobility metrics.

DOI 10.5162/iCCC2026/2.3

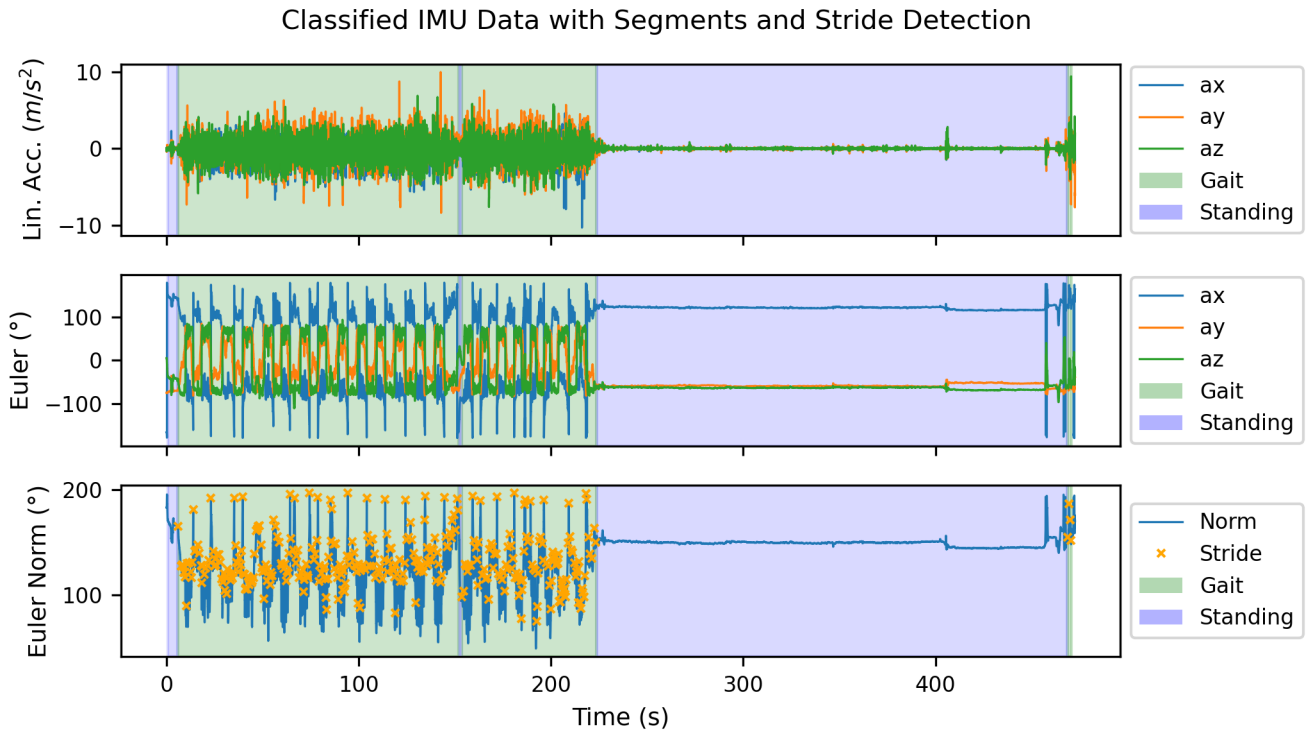


Fig. 2. The figure displays (top to bottom): Linear Acceleration (ax,ay,az in m/s2), Euler Angles (yaw,pitch,roll in degrees), and the Euler Norm (in degrees). The background color indicates the activity class predicted by the 1D-CRNN: Gait (green) and Standing (purple). In the bottom plot, the Euler Norm is used for step detection; the orange 'x' markers indicate the detected steps (or strides) within the segmented gait periods. The model successfully isolates periods of continuous walking from quiescent standing.

The method is lightweight, relies on highly acceptable, unobtrusive textile sensing, and is thus well-suited for large-scale or near-real-time monitoring applications, particularly in elderly populations.

Future work will focus on:

- Extending the activity classification to include a full spectrum of Activities of Daily Living (ADL).
- Validating the pipeline's performance and the accuracy of the derived gait parameters against a gold standard in the target geriatric cohort.
- Refining step detection and stride estimation through anthropometric calibration or sensor fusion techniques.

Funding

This research was funded by German Federal Ministry of Education and Research grant number 03WIR6308D. The APC was funded by Projektträger Jülich.

Acknowledgments

We would like to thank the company MedConNet GmbH, FiberCheck GmbH, Sächsisches Textilforschungsinstitut e.V. and Smart Textiles Hub GmbH for developing together this prototype.

References

[1] Studenski, S., Perera, S., Patel, K., Rosano, C., Faulkner, K., Inzitari, M., ... & Guralnik, J. (2011). Gait speed and survival in older adults. Jama, 305(1), 50-58.

[2] Verghese, J., Wang, C., Lipton, R. B., Holtzer, R., & Xue, X. (2007). Quantitative gait dysfunction and risk of cognitive decline and dementia. Journal of Neurology, Neurosurgery & Psychiatry, 78(9), 929-935.

[3] Hausdorff, J. M., Rios, D. A., & Edelberg, H. K. (2001). Gait variability and fall risk in community-living older adults: a 1-year prospective study. Archives of physical medicine and rehabilitation, 82(8), 1050-1056.

[4] Shin, J., Kweon, H. J., & Choi, J. (2025). Assessment of Gait Parameters Using Wearable Sensors and Their Association With Muscle Mass, Strength, and Physical Performance in Korean Older Adults: Cross-Sectional Study. JMIR Formative Research, 9, e63928.

[5] Voisard, C., De l'Escalopier, N., Ricard, D., & Oudre, L. (2024). Automatic gait events detection with inertial measurement units: healthy subjects and moderate to severe impaired patients. Journal of NeuroEngineering and Rehabilitation, 21(1), 104.

[6] Stenger, R., Hozhabr Pour, H., Teich, J., Hein, A., & Fudickar, S. (2025). Gait Event Detection and Gait Parameter Estimation from a Single Waist-Worn IMU Sensor. Sensors, 25(20), 6463.

[7] CEVA, Inc. (2021): BNO080/BNO085 Intelligent 9-Axis Absolute Orientation Sensor – Datasheet, Rev. 1.11. CEVA Technologies Ltd., Hillcrest Labs Division. Available online: https://www.ceva-ip.com/wp-content/uploads/BNO080_085-Datasheet.pdf(accessed on 13 October 2025).

[8] Steinmetzer, T., Wieczorek, F., Naake, A., Wolf, P., Braun, A., & Michel, S. (2025). Smart e-Textile Singlet Prototype and Concept: Multi Sensor Sensing for Geriatric Monitoring. Bioengineering, 12(11), 1275.

[9] Steinmetzer, T., & Michel, S. (2025). Towards Wearable Respiration Monitoring: 1D-CRNN-Based Breathing Detection in Smart Textiles. Sensors, 25(22), 6832.

DOI 10.5162/iCCC2026/2.4

Implantable MEMS Sensor Platform for Measuring Heart Activity

Sören Schäfer[1*], Peter Detemple[1], Max Haberbusch[2], Oliver Haverbeck[1], Juri Magomajew[1], Francesco Moscato[2], Stefan Schmitt[1]
[1]: Fraunhofer Institute for Microengineering and Microsystems IMM, Mainz, Germany
[2]: Center for Medical Physics and Biomedical Engineering, Medical University of Vienna, Vienna, Austria
[*]soeren.schaefer@imm.fraunhofer.de

Abstract

Heart transplants usually require the severing of the vagus nerve, which connects the heart muscle to the central nervous system. Consequently, the parasympathetic control of the heart, which is responsible for lowering heart activity during periods of relaxation, is impaired. In addition to limitations in the everyday lives of patients, this causes, for example, a permanently elevated resting heart rate and thus accelerated aging of the tissue. One possible solution for restoring the vagal control is to externally stimulate the severed nerve pathway via an implantable electrode array. The response of the heart muscle to a specific stimulation signal must then be detected and ideally be fed back to the stimulation unit.
We report on an implantable sensor platform that is developed for continuous detection of the heart activity. Implanted devices are an attractive alternative to non-invasive measurement methods in terms of reliability and comfort. The MEMS platform has three axially arranged piezoresistive force sensors on a glass substrate, which are encapsulated with a biocompatible material. Ex-vivo tests in a pressure chamber are followed by in-vivo tests on pig hearts, in which cardiac activity is controlled by administering dobutamine. In-vivo stability tests show a reliable measurement signal even after one month and underline the biocompatibility.

1 Introduction

A heart transplant (approximately 3,500 per year worldwide) is the only medical treatment option for patients with end-stage heart failure [1, 2]. However, the denervation of the heart caused by the operation means that the controlled increase and decrease in heart rate can no longer be fully guaranteed (chronotropic incompetence). For patients, this results in limitations in physical endurance and typically an increased resting heart rate. The latter in particular leads to accelerated aging or hardening of the heart muscle tissue (fibrosis) due to permanently increased contraction forces, which usually necessitates another transplant after a few years.

One solution could be targeted stimulation of the vagus nerve connection, e.g., via an implantable electrode interface [3]. The vagus nerve is a central component of the parasympathetic nervous system, which is responsible for rest and relaxation, including the regulation of heart rate. In order to close the feedback loop, in addition to stimulating the nerve, the heart activity must be measured, processed, and evaluated as a response and then fed back to the interface in the next step.

The article describes the development of an implantable MEMS sensor platform that continuously detects cardiac activity using pressure sensors. The challenge here lies in the biocompatible encapsulation of commercial MEMS pressure sensors and the robust electrical signal readout, so that cardiac activity can be measured over the long term without impairing the function of the heart itself. The article is divided into the design of the platform and subsequent ex-vivo and in-vivo tests in a pig model.

2 Concept development

2.1 Approach

Figure 1a) shows an oblique view of the 3D model of an early concept design of the sensor platform. Three pressure sensors are positioned on a rigid substrate, e.g., a glass plate. The entire arrangement, including electrical connections, is then encapsulated with a biocompatible material, e.g., silicone. Figure 1b) schematically shows how the sensor platform is implanted in the heart muscle and illustrates the forces acting on it. The special geometry, shown here as a hillock structure in the encapsulation, ensures that two of the three pressure sensors (S1 and S3) experience a higher force than S2 (F_1, $F_3 > F_2$). In addition to redundancy and an extended measuring range, this should ideally also provide information about the stiffness of the heart muscle tissue, provided that this has an influence on the difference between the respective signals output of S2 over S1 and S3.

2.2 Sensor design

The starting point for implementation is the question of the required dimensions of the platform (maximum sensitivity vs. minimum invasiveness) and the selection of suitable sensors and materials. To estimate the expected forces, analytical models of the stress on the left ventricle are used [4, 5]. Since the encapsulation material dampens the force acting on the sensors, the arrangement shown in Fig. 1b) is examined in an FEM simulation using suitable assumptions for the elasticity modulus of the materials involved. Overall, a value between 12 and 48 kPa is expected for the normal component of the stress. Commercially available

DOI 10.5162/iCCC2026/2.4

piezoresistive elements (SM5108E, Silicon Microstructures, Inc.) were used as bare dies for the sensors. These measure the absolute pressure with a sensitivity of 0.6 mV/kPa and come with a measuring range from 0 to approx. 200 kPa.

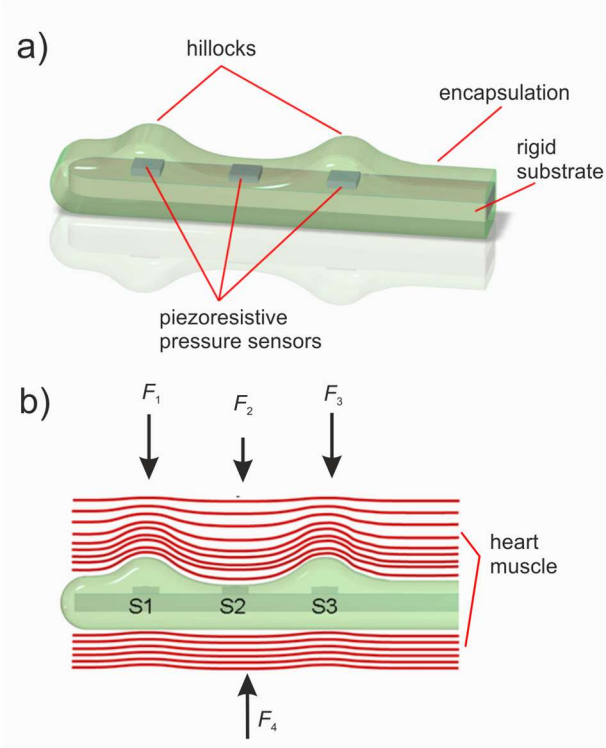


Figure 1 a) 3D model of the sensor concept with three piezoresistive pressure sensors, b) different forces F_1 through F_4 on the sensor platform implanted in the heart muscle.

3 Implementation

Figure 2 shows a 3D model as well as photographs of an unencapsulated sensor platform. A glass plate with dimensions of (27 × 2) mm^2 with lithographically structured Au layers serves as rigid substrate. The deposition of Au layers is in principle a two-step process, starting with a lift-off structuring process for an e-beam evaporated 200 nm-thick layer of Au with Cr/Ni adhesion layers. This serves as a starting layer for an electroplating step during which the Au layer gets thickened up to 3 µm. Afterwards the sensor elements are fixated via epoxy adhesive and wire-bonded to the respective base pads on the substrate. The two outer sensor elements S1 and S3 are positioned on pillars of silicon with the same base area as the sensor element and a height of, here, 3 mm.

The encapsulation process must on the one hand ensure a reliable sealing of the electronics and on the other hand must leave the delicate wire bond contacts intact. It is furthermore important to have a defined thickness of the encapsulation material above the pressure sensors to ensure reproducible results.

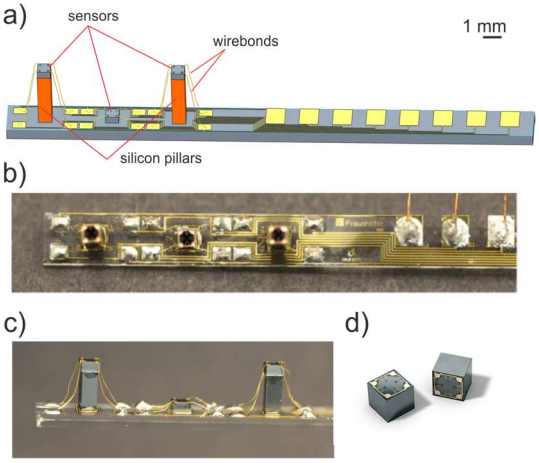


Figure 2 a) 3D model of the unencapsulated sensor platform with sensor elements on pillars of silicon. b), c) electrically connected unencapsulated sensor platform on a glass substrate from top and front view, respectively. d) Image of bare die pressure sensors with dimensions of (0.69 × 0.69 × 0.40) mm^3.

This has been realized by a 3D-printed casting mold that has been modeled based on the topography of the sensor elements as shown in Figs. 3a), b). Once the sensor plate is placed within the mold parts, liquid silicone (Dragonskin) is filled in via the tiny hole in part 3 and left for curing. Figure 3c) shows photographs of the encapsulated front side after demounting from the mold.

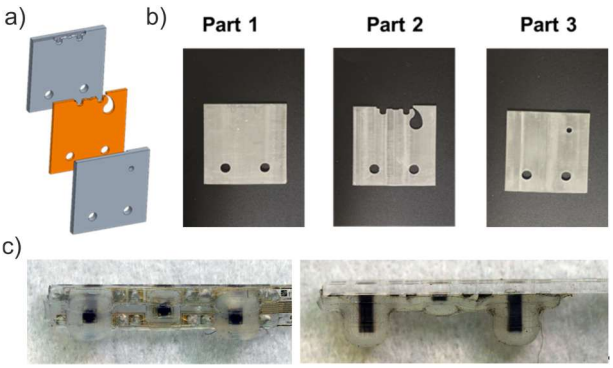


Figure 3 a) 3D model of the three part casting mold. b) Finished 3D-printed mold parts. Silicone is filled in via the tiny hole in part 3. c) top and side views of the encapsulated sensor platform.

DOI 10.5162/iCCC2026/2.4

4 Functional tests

4.1 Ex-vivo

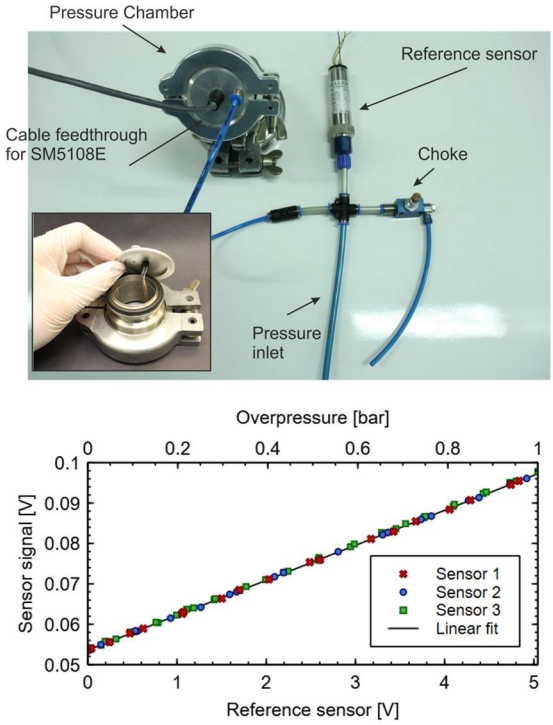


Figure 4 Top shows a photograph of the air-tight pressure chamber containing the connected sensor platform as indicated in the inset. A reference pressure sensor is connected in parallel. Bottom shows the pressure signal of each of the three sensors over the voltage of the reference as well as a linear fit (black line). The markers for each line are reduced for visibility.

First, each encapsulated sensor platform is placed in an in-house built pressure chamber and is tested as shown in Fig. 4. The chamber comes with a cable feedthrough to extract the signal of each of the three sensors. A reference sensor (Omega Keller, type 21S/80549.3) is connected in parallel. The reference sensor and the SM5108E of the platform show 0 V and 53 mV for atmospheric pressure and 5 V and 97 mV for 1 bar overpressure, respectively. All sensors show a linear behavior versus the overpressure. The linear fit yields a slope of 8.6 mV/V which corresponds to a sensitivity of about 0.43 mV/kPa. This is slightly less than the theoretical sensitivity given by the manufacturer but plausible, given that the sensors are encapsulated by a layer of silicone.

In addition, the sensor platform was tested by pressing it into a soft tissue like a block of gelatin or pieces of pork loin (not shown here). The main insight from these tests is that the height differences of S1 and S3 over S2 as shown in Figs. 2 and 3 is too large for S2 to give out a signal. Hence, the height discrepancy was reduced to 200 µm by mounting S2 on a Si pillar as well.

4.2 In-vivo.

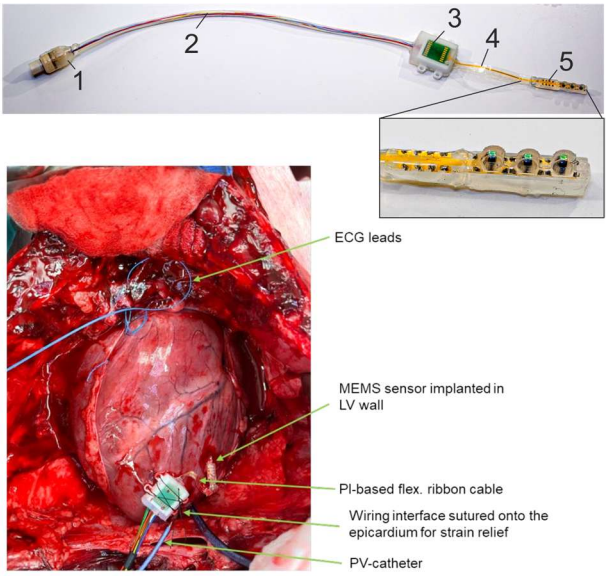


Figure 5 Top photograph shows the device prepared for in-vivo testing: connector (1), wires guided through silicone tubing (2), ASIC for AD conversion (3), custom-made flexible conductor ribbon (4), encapsulated sensor platform (5). Bottom photograph shows the device implanted in the left ventricular wall of a pig heart (courtesy Medical University of Vienna).

After ex-vivo testing, the sensor device is further prepared for in-vivo testing in a pig heart as shown in Fig. 5. The glass substrate is further encapsulated in a thin layer of silicone for protection. The connection to the ADC ASIC [6] is over a custom-made, PI-based flexible ribbon cable for strain relief. Figure 5 additionally shows the in-vivo application in a pigs heart. Here, the sensor array is placed in the left ventricular wall. However, also other positions of the sensors were tested with respect to responsivity.

Figure 6 exemplarily shows the sensor signals as well as electrocardiogram signal (ECG) for two different stages: a) refers to pre and b) to post injection of dobutamine, which stimulates heart activity. The signal level is reduced compared to calibration obtained during calibration in Fig. 4 which is likely due to further encapsulation and ohmic losses. However, the sensors clearly respond to the forces during the cardiac cycle and reflect the heart activity. The instantaneous heart rate can be easily extracted. Other than that, the signal form changes upon administration of dobutamine and indicates a change in contractility. The long term stability and biocompatibility have been evaluated and are described in detail in Ref. [7]. After a 30-day period the signal amplitude of the sensors decreases. However, heart rate and certain signatures of the signal form can still be extracted. Even after 120 days of implantation, no severe immune response has been detected which underlines the biocompatibility.

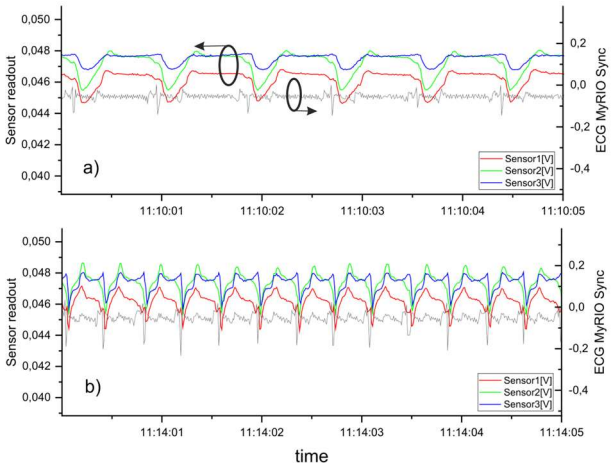


Figure 6 Readout signals of the three sensors in mV during in-vivo testing in a pigs heart. The grey curves indicate the ECG signal. a) and b) refer to before and after injection of dobutamine, respectively.

5 Acknowledgments

We acknowledge the European Commission for funding the H2020-FETPROACT-2018-2020 NEUHEART Project No. 824071, which made this work possible. The authors furthermore thank the collaborators from Sant'Anna University in Pisa, Medizinische Universität Wien and University College London that helped acquiring parts of the shown results. Sören Schäfer. acknowledges funding by the German Federal Ministry of Education and Research in the context of the federal-state program "FH-Personal" under the grant number 03FHP147A (REQUAS).

6 References

[1] E. Y. Birati and J. E. Rame, "Post-heart transplant complications", *Critical care clinics*, vol. 30, no. 3, pp. 629–637, 2014, doi: 10.1016/j.ccc.2014.03.005.

[2] A. Jaiswal *et al.,* "Clinical outcomes of older adults listed for heart transplantation in the United States", *Journal of the American Geriatrics Society*, vol. 69, no. 9, pp. 2507–2517, 2021, doi: 10.1111/jgs.17271.

[3] M. Capogrosso *et al.,* "A brain–spine interface alleviating gait deficits after spinal cord injury in primates" (in En;en), *Nature*, vol. 539, no. 7628, pp. 284–288, 2016, doi: 10.1038/nature20118.

[4] P. Gould, D. Ghista, L. Brombolich, and I. Mirsky, "In vivo stresses in the human left ventricular wall: analysis accounting for the irregular 3-dimensional geometry and comparison with idealised geometry analyses", *Journal of Biomechanics*, vol. 5, no. 5, pp. 521–539, 1972, doi: 10.1016/0021-9290(72)90009-7.

[5] D. N. Ghista and H. Sandler, "An analytic elastic-viscoelastic model for the shape and the forces in the left ventricle", *Journal of Biomechanics*, vol. 2, no. 1, pp. 35–47, 1969, doi: 10.1016/0021-9290(69)90040-2.

[6] N. Neshatvar, M. Schormans, D. Jiang, S. Schmitt, P. Detemple, and A. Demosthenous, "An Implantable Phase Locked Loop MEMS-Based Readout System for Heart Transplantation", *IEEE Trans. Circuits Syst. II*, vol. 69, no. 10, pp. 4168–4172, 2022, doi: 10.1109/TCSII.2022.3190796.

[7] C. Zinno *et al.,* "Implementation of an epicardial implantable MEMS sensor for continuous and real-time postoperative assessment of left ventricular activity in adult minipigs over a short- and long-term period", *APL bioengineering*, vol. 8, no. 2, p. 26102, 2024, doi: 10.1063/5.0169207.

DOI 10.5162/iCCC2026/2.5

Trainingssteuerung im Leistungssport: Integration multimodaler Daten durch Datenraumtechnologien

Training control in competitive sports: Integration of multimodal data through data space technologies

Anja Burmann, Fraunhofer-Institut für Software- und Systemtechnik ISST, Dortmund, Deutschland
anja.burmann@isst.fraunhofer.de
Sebastian Schröder, Enari GmbH, Garching bei München, Deutschland
Florian Lauf, Fraunhofer-Institut für Software- und Systemtechnik ISST, Dortmund, Deutschland

Kurzfassung

Die elektronische Sportakte (eSA) stellt einen innovativen Ansatz dar, der leistungsdiagnostische, sportmedizinische und trainingsbezogene Daten im Leistungssport über einen föderierten Datenraum miteinander verbindet. Durch die Kombination aus Datensouveränität, semantischer Interoperabilität, sicheren Verarbeitungsumgebungen und neutraler Governance entsteht eine Infrastruktur, die sowohl Athlet:innen als auch Trainer:innen, medizinischen Einrichtungen und der Forschung zugutekommt. Moderne Datenraumarchitekturen bieten somit eine tragfähige Grundlage für einen digital unterstützten, evidenzbasierten und verantwortungsvollen Leistungssport.

Abstract

The electronic athlete file (eSA) represents an innovative approach that combines performance diagnostics, sports medicine, and training-related data in competitive sports via a federated data space. The combination of data sovereignty, semantic interoperability, secure processing environments, and neutral governance creates an infrastructure that benefits athletes, coaches, medical facilities, and research. Modern data space architectures thus provide a solid foundation for digitally supported, evidence-based, and responsible competitive sports.

1 Einleitung

1.1 Motivation

Die Digitalisierung des Leistungssports hat in den vergangenen Jahren zu einer erheblichen Zunahme an Mess- und Gesundheitstechnologien geführt. Wearables, Diagnostikgeräte, Trainingssoftware und sportmedizinische Dokumentationssysteme generieren eine Vielzahl multimodaler Daten, die potenziell wertvolle Erkenntnisse zur Belastungssteuerung, Leistungsentwicklung und Prävention liefern können. In der Praxis sind diese Daten jedoch in proprietären Silos eingeschlossen und kaum organisationsübergreifend zugänglich. Dadurch manifestieren sich Informationsbrüche, die das Potenzial integrierter sportwissenschaftlicher Entscheidungen zwischen Trainer:innen, Diagnostiker:innen, Mediziner:innen und Forschenden nicht ausschöpfen [1].

Zentrale Gründe für diese Fragmentierung liegen in einem historisch gewachsenen Ökosystem von Akteursgruppen, Softwaresystemen, und zunehmend multimodale und mobile Sensortechnik. Daraus ergibt sich eine fehlende Systemarchitektur, fehlende Interoperabilität und semantische Standardisierung, sowie organisationsübergreifender Schnittstellen und Workflows. Datenraumarchitekturen bieten einen Lösungsweg, um föderierte und souveränitätsorientierte Datennutzung domänenübergreifend zu ermöglichen [2, 3]. Auch im Gesundheitsbereich wird dieser Ansatz zunehmend relevant, wie der European Health Data Space (EHDS) zeigt, der föderierte Verarbeitung, interoperable Datenmodelle und sichere Datenverarbeitungsumgebungen (Secure Processing Environments, SPE) vorsieht [4].

Vor diesem Hintergrund untersucht dieser Beitrag die konzeptionelle und technische Ausgestaltung einer elektronischen Sportakte (eSA), welche als Applikation auf Basis von Datenraumtechnologien verteilte Datenquellen für leistungsdiagnostische, sportmedizinische und trainingsbezogene Daten integriert. Der Fokus liegt dabei auf der Frage, wie sportbezogene Gesundheits- und Leistungsdaten organisationsübergreifend, dezentral nutzbar gemacht werden können, und dabei gleichzeitig die Datensouveränität der Athlet:innen in den Fokus zu stellen. Der Beitrag analysiert bestehende Herausforderungen im Leistungssport, leitet Anforderungen an Interoperabilität, Datensouveränität und Governance ab und transferiert diese auf eine Datenraumarchitektur.

2 Hintergrund

2.1 Informationstechnologie im Leistungssport

Der Leistungssport ist durch eine dezentrale Organisationsstruktur gekennzeichnet, in der Trainingsgruppen, Verbände, Olympiastützpunkte, Vereine und externe Diagnostikzentren zusammenwirken, aber unterschiedliche digitale Infrastrukturen verwenden. Weiterhin werden zunehmend

DOI 10.5162/iCCC2026/2.5

mobile Sensorik und körpernahe Systeme zum Monitoring eingesetzt, um Daten über den Gesundheits- und Trainingszustand der Athlet:innen zu erheben. Dadurch entsteht eine multimodale Datenlandschaft, die in proprietären Systemen gehalten werden und einem spitzen Nutzungszweck dienen [6].
Zudem führt die zunehmende Verbreitung von Wearables im Spitzen- wie im Breitensport zu einer massiven Ausweitung von Datenvolumen, -geschwindigkeit und -vielfalt, die durch klassische Analyseverfahren kaum mehr zu bewältigen ist. Studien aus der Ganganalyse zeigen, dass moderne Sensorsysteme, wie etwa IMUs, 3D-Motion-Capture oder multimodale Messplattformen pro Athlet:in Hunderttausend bis Millionen Zeitreihenpunkte erzeugen und damit die Kriterien von Big Data erfüllen (Volume, Velocity, Variety, Veracity, Value) [7, 8].
Auch in der sportartspezifischen Leistungsdiagnostik steigen die Anforderungen: Wearables liefern in Echtzeit kinematische, dynamische und physiologische Parameter, ermöglichen Feldmessungen unter Wettkampfbedingungen und erzeugen hochfrequente, heterogene und zum Teil fehlerbehaftete Datenströme, deren Qualität von Sensorfunktion, Kalibrierung und Trägerartefakten abhängt [9].
Parallel zeigt sich in aktuellen Übersichtsarbeiten, dass Wearables zunehmend in komplexe Ökosysteme eingebettet werden, die sowohl Trainingssteuerung, Gesundheitsmonitoring als auch taktische und biomechanische Analysen unterstützen. Dabei wird auf Chancen wie kontinuierliches Monitoring, Echtzeit-Feedback, Individualisierung des Trainings und Integration in digitale Trainingsplattformen verwiesen. Risiken wie mangelnde Standardisierung, Datenschutzprobleme, unzureichende Validierung vieler Systeme sowie Fehlinterpretationen durch Nutzer:innen dürfen dabei nicht außen vor gelassen werden [10].
Diese technologisch getriebene Entwicklung verstärkt die Fragmentierung der Datenlandschaft, da verschiedene Stakeholder, wie Trainer:innen, Wissenschaft, Diagnostikzentren, Athlet:innen, Verbände jeweils proprietäre Systeme nutzen, unterschiedliche Datenformate erzeugen und selten gemeinsame Schnittstellen existieren. Damit wächst der Bedarf an interoperablen Datenräumen, standardisierten Austauschprozessen, gemeinsamen Datenmodellen und semantischen Ebenen, um die Vielzahl dezentral erzeugter sportwissenschaftlicher Daten integrierbar, vergleichbar und auswertbar zu machen.

2.2 Konzept des EHDS

Im Kontext der Gesundheitsversorgung in Europa entwickelt sich mit dem European Health Data Space (EHDS) eine internationale Gesetzgebung, die eine Grundlage für eine organisationsübergreifende Datenverfügbarkeit zum Zwecke der Versorgung (primäre Datennutzung) und Forschung (sekundäre Datennutzung) schafft. Hierfür definiert der EHDS Interoperabilitätsanforderungen an Health-IT-Systemhersteller, spezifiziert Infrastrukturen für die beiden Zieldimensionen, definiert föderierte, sichere Verarbeitungsumgebungen, und schreibt semantische Gesundheitsdatenstandards vor [4]. Von diesen Prinzipien kann bei konzeptioneller Übertragung auch der Leistungssport profitieren, da sportmedizinische und leistungsdiagnostische Daten ebenso zu primären Zwecken der Trainingssteuerung und sekundär für Forschung und Entwicklung organisationsübergreifend genutzt werden sollen [11].

2.3 Datenraumtechnologien als Infrastrukturbasis

Datenräume stellen eine dezentrale Infrastrukturform dar, die den Austausch und die Nutzung von Daten unter strikten Bedingungen der Datensouveränität ermöglichen. Im Zentrum steht das Konzept der föderierten Datenhaltung: Daten verbleiben bei den ursprünglichen Datenhaltern, und werden über Konnektoren, Metadatenkataloge, Identitäts- und Zugriffsmanagement sowie Nutzungsrichtlinien so verfügbar gemacht, dass sie in organisationsübergreifenden Anwendungsszenarien genutzt werden können [12, 13]. Die International Data Spaces Reference Architecture (IDSA-RAM) beschreibt hierfür ein Set aus Governance- und Technologiebausteinen, die transparente und auditierbare Datennutzung sicherstellen [2].

2.4 Datenrauminitiative sphin-X

Die Ausgestaltung des EHDS bedarf kanalisierter Beiträge aus allen Organisationen der Gesundheitsbranche, und damit auch der Gesundheitswirtschaft. Die Initiative sphin-X e.V. verfolgt das Ziel, einen interoperablen und souveränitätswahrenden Datenraum anhand von etablierten Referenzarchitekturen für Akteure des Gesundheitswesens und der Gesundheitswirtschaft aufzubauen, und einen wichtigen Beitrag für den EHDS zu leisten [14]. Die Initiative entwickelt technische und organisatorische Bausteine – darunter föderierte Datenzugänge, standardisierte Schnittstellenprofile, sichere Verarbeitungsumgebungen und ein domänenspezifisches Governance-Modell. Neben diesen Bausteinen schafft der sphin-X e.V. eine neutrale Entwicklungsumgebung für die beteiligten Akteure am Aufbau einer eSA innerhalb des Gesundheitsdatenökosystems.

3 Anwendungsfall eSA im föderierten Datenraum

3.1 Primäre Datennutzung: Leistungssteuerung

Der Leistungssport ist ein Paradebeispiel für multimodale und multidisziplinäre Datenprozesse. Leistungsdiagnostische Daten wie Laktatschwellen, VO_2max, Kraftverläufe oder Bewegungsanalysen, werden typischerweise in spezialisierten Zentren erhoben und verbleiben dort. Wearables liefern hochfrequente Zeitreihen zu Belastung und Regeneration, während sportmedizinische Daten in Arztpraxen oder physiotherapeutischen Einrichtungen entstehen. Für Trainer:innen sind diese Daten jedoch nur dann wertvoll, wenn sie im Kontext von Trainingsplänen, subjektiven Belastungswahrnehmungen und langfristigen Entwicklungsverläufen betrachtet werden können.
Voraussetzung für eine sinnvolle organisationsübergreifende Nutzung ist, dass diagnostische Befunde, Wearable-

Daten und subjektive Ratings mit einheitlichen Protokollen erhoben, annotiert und versioniert werden (z. B. standardisierte Testprotokolle, Erfassungszeitpunkte, Metadaten zu Sensor-typen und Kalibrierung).
Die eSA ermöglicht diesen organisationsübergreifenden Kontext, indem relevante Daten über föderierte Zugriffe temporär zusammengeführt und visualisiert werden. Ein Trainer kann so beispielsweise Belastungsdaten der letzten Wochen mit diagnostischen Schwellenwerten und aktuellen medizinischen Empfehlungen abgleichen oder in Kontext setzen. Diagnostiker:innen können Vorwerte heranziehen, um Abweichungen präziser einzuordnen, während Mediziner:innen Risiken wie Überlastungen oder kardiale Auffälligkeiten besser erkennen können. Ohne diese Kontextinformationen bleiben Werte wie VO_2max oder Belastungsindizes aus Wearables nur bedingt vergleichbar und können Trainingsentscheidungen sogar verzerren. Die eSA kann darüber hinaus nicht nur historische Verläufe visualisieren, sondern auch regelbasierte oder modellgestützte Benachrichtigungen (z. B. plötzliche Abweichungen von Belastungs-/Regenerationsprofilen) bereitstellen. Diese werden explizit als Entscheidungsunterstützung und nicht als automatisierte Entscheidung implementiert. Zentral bei all diesen Punkten ist die Einbindung der Athlet:innen: Über ein Consent- und Transparenz-Cockpit innerhalb der eSA können sie nachvollziehen, welche Daten in welchen Kontexten (Training, Medizin, Forschung) genutzt werden und Zugriffe situativ erlauben oder einschränken.

3.2 Sekundäre Datennutzung: Forschung und Entwicklung

Daten aus der Leistungssteuerung können entlang der Prinzipien des EHDS über Datenraumtechnologie für Sekundärnutzung erschlossen werden. Durch die Nutzbarmachung pseudonymisierter Daten über sichere Verarbeitungsumgebungen kann sportwissenschaftliche Forschung ohne zusätzliche Belastung der Athlet:innen durchgeführt werden. Algorithmen zur Belastungssteuerung oder Verletzungsprävention können direkt in diesen Umgebungen ausgeführt werden, ohne dass Rohdaten die Systeme der Datenhaltenden verlassen müssen. Die eSA integriert damit die Datennutzung von der Erhebung über Analyse und Entscheidung bis zur Sekundärnutzung mit dem Beitrag von aggregierten Informationen zu einem wachsenden Wissensschatz. Im Spitzensport ist aufgrund kleiner Kollektive und charakteristischer Belastungs- und Leistungsprofile selbst bei Pseudonymisierung ein erhöhtes Re-Identifikationsrisiko gegeben. Die eSA begegnet diesem Risiko durch Maßnahmen wie strenge Aggregationsgrade, Zugriffsgenehmigungen über Data-Access-Mechanismen und die Trennung von Identitäts- und Nutzungsdaten. Ein Governance-Rahmen stellt sicher, dass aus Sekundäranalysen erzielte Erkenntnisse nicht nur kommerziell verwertet, sondern auch in Form von Leitlinien, Feedback oder innovativen Services an Athlet:innen und betreuende Organisationen zurückgespielt werden. Neben der semantischen Standardisierung adressiert die eSA auch technische (Schnittstellen, Protokolle) und organisatorische Interoperabilität (Prozessharmonisierung, Zuständigkeiten), um föderierte Datenflüsse praktikabel zu machen.
Das zeigt, dass die heterogene Datenlandschaft im Bereich der Leistungssteuerung sich besonders gut für einen modularen, föderierten Architekturansatz eignet. Durch Standardisierung auf semantischer Ebene, verteilte Verarbeitung und rollenbasierte Zugriffe können Erkenntnisse für Training, Regeneration und Reha effizienter gewonnen werden, ohne zentrale Datenspeicher aufzubauen.

4 Architekturbasis der eSA

Die technische Architektur der eSA besteht aus mehreren interoperablen Komponenten, die gemeinsam einen föderierten Datenraum bilden. Jede teilnehmende Organisation, d.h. Datengebende und Datennutzende, betreibt einen Konnektor, der als geschütztes Gateway fungiert und Datenzugriffe kontrolliert. Dieser Konnektor setzt die Nutzungsregeln um, die in den Consent-Vereinbarungen der Athlet:innen definiert sind, und stellt sicher, dass Daten nur in zulässigen Szenarien geteilt oder verarbeitet werden.
Ein Metadatenkatalog dient der Auffindbarkeit von Datenquellen und beschreibt, unter welchen Bedingungen diese genutzt werden dürfen. Zur semantischen Interoperabilität werden offene Standards eingesetzt, insbesondere HL7 FHIR für gesundheitsnahe Informationen sowie standardisierte Austauschformate für sportdiagnostische Messdaten und Sensorik. Jede Ressource erhält eine eindeutige URI, sodass Daten virtualisiert verknüpft werden können, auch wenn sie physisch verteilt gespeichert sind.
Zur sekundären Datennutzung kommen sichere Verarbeitungsumgebungen (engl. Secure Processing Environments, SPE) zum Einsatz, in denen statistische oder KI-basierte Analysen unter kontrollierten Bedingungen durchgeführt werden können. Dieser Mechanismus, der in Datenräumen als Compute-to-Data bezeichnet wird, gewährleistet hohe Datenschutzstandards und ermöglicht gleichzeitig evidenzbasierte Weiterentwicklung von Trainings- und Präventionsstrategien. Die gesamte Architektur wird durch Governance-Regeln umrahmt, die Transparenz, Fairness und vertrauenswürdige Zusammenarbeit sichern.
Die Entwicklung und Moderation dieser technischen, organisatorischen und ethischen Rahmenbedingungen wird durch die neutrale Trägerstruktur des sphin-X e.V. begleitet. Eine solche Instanz fungiert als Entwicklungsraum für den vorwettbewerblichen Technologiereifegrad, und muss perspektivisch in eine Governance-Autorität überführt werden, die Zertifizierungsverfahren, Rollenmodelle, Interoperabilitätsstandards und Auditprozesse koordiniert.
Im Falle der eSA kann eine solche Rolle von einer gemeinwohlorientierten Institution übernommen werden, die Innovation, Transparenz und Stabilität im Ökosystem sicherstellt. Diese Instanz ermöglicht den Austausch zwischen Sportpraxis, Wissenschaft und Technologieentwicklung.
Das konzeptionelle Datenmodell der eSA orientiert sich an den Interoperabilitäts- und Governance-Prinzipien des EHDS und integriert semantische Interoperabilität, Zweck-

bindung und Datensouveränität. Es modelliert sportbezogene Gesundheits- und Leistungsdaten als semantisch verknüpfte Entitäten entlang der Ebenen Datenerhebung (Mess- und Diagnosedaten), Kontextualisierung (Trainingsereignisse, medizinische Zustände) und Ableitung (aggregierte Kennzahlen, Entscheidungsunterstützung).

5 Diskussion

5.1 Herausforderungen

Die eSA zeigt, dass föderierte Datenraumarchitekturen einen geeigneten Rahmen für die Integration sportmedizinischer und leistungsdiagnostischer Daten bieten. Dennoch bestehen Herausforderungen, wie die Akzeptanz bei Trainer:innen und Athlet:innen, die Beteiligung von Systemherstellern und die Entwicklung einer langfristigen Betriebs- bzw. Governance-Struktur. Die Akzeptanz der beteiligten Akteure hängt stark von der Benutzerfreundlichkeit der Systeme, den Mehrwertdiensten und der Transparenz der Governance ab. Technisch ist die Integration heterogener Systeme machbar und erfordert klare Schnittstellenstandards. Organisatorisch müssen Verantwortlichkeiten für Betrieb, Zertifizierung und Monitoring eindeutig definiert werden.

5.2 Chancen

Trotz dieser Herausforderungen bietet die eSA die Chance, den Leistungssport datenbasiert, sicher und nachhaltig weiterzuentwickeln. Sie trägt dazu bei, multimodale Daten nahe Echtzeit für die personalisierte Trainingssteuerung nutzen zu können, Doppeluntersuchungen zu vermeiden, Risiken frühzeitig zu erkennen und wissenschaftliche Evidenz stärker in den Trainingsalltag zu integrieren.

5.3 Fazit

Die eSA beschreibt den Ansatz, leistungsdiagnostische, sportmedizinische und trainingsbezogene Daten im Leistungssport über einen föderierten Datenraum miteinander zu verbinden. Datenraumtechnologie definiert eine flexible Schnittstelle und Nutzungsbedingungen, sodass dezentrale Datenquellen im Ökosystem integriert werden können. Durch die Kombination aus Datensouveränität, semantischer Interoperabilität, sicheren Verarbeitungsumgebungen und neutraler Governance entsteht eine Infrastruktur, die sowohl Athlet:innen als auch Trainer:innen, medizinischen Einrichtungen und der Forschung zugutekommt. Moderne Datenraumarchitekturen bieten somit eine tragfähige Grundlage für einen digital unterstützten, evidenzbasierten und verantwortungsvollen Leistungssport.

6 Literatur

[1] Torres-Ronda, L.; Beanland, E.; Whitehead, S. *et al.* Tracking Systems in Team Sports: A Narrative Review of Applications of the Data and Sport Specific Analysis. *Sports Med - Open* **8**, 15 (2022). https://doi.org/10.1186/s40798-022-00408-z

[2] Steinbuss, S.; Otto, B.; Teuscher, A.; Pettenpohl, H.: IDS Reference Architecture Model 4 (IDS-RAM 4). (2022).

[3] Gaia-X AISBL: Gaia-X – Architecture Document - 22.04 Release, 2022.

[4] European Commission: European Health Data Space Regulation. 2025.

[5] Bundesinstitut für Sportwissenschaft: Chancen neuer (techno-logischer) Entwicklungen. https://www.bisp.de/DE/Forschungsschwerpunkte/TechnologischeEntwicklung/technologische-entwicklung_node.html. [abgerufen 10.12.2025]

[6] Montull, L.; Slapšinskaitė-Dackevičienė, A.; Kiely, J. *et al.* Integrative Proposals of Sports Monitoring: Subjective Outperforms Objective Monitoring. *Sports Med - Open* **8**, 41 (2022). https://doi.org/10.1186/s40798-022-00432-z

[7] Phinyomark, A.; Petri, G.; Ibáñez-Marcelo, E. *et al.* Analysis of Big Data in Gait Biomechanics: Current Trends and Future Directions. *J. Med. Biol. Eng.* **38**, 244–260 (2018). https://doi.org/10.1007/s40846-017-0297-2

[8] Phinyomark, A.; Petri, G.; Ibáñez-Marcelo, E.; Osis, S. T., & Ferber, R. (2018). Analysis of Big Data in Gait Biomechanics: Current Trends and Future Directions. *Journal of medical and biological engineering*, *38*(2), 244–260. https://doi.org/10.1007/s40846-017-0297-2

[9] Camomilla, V.; Bergamini, E.; Fantozzi, S.; Vannozzi, G. Trends Supporting the In-Field Use of Wearable Inertial Sensors for Sport Performance Evaluation: A Systematic Review. *Sensors* **2018**, *18*, 873. https://doi.org/10.3390/s18030873

[10] Drummond, K.; Lewandowski, B.; Massaroni, C.; (2022) Editorial: Wearable Technology for Human Performance. *Front. Physiol.* 13:871159. doi: 10.3389/fphys.2022.871159

[11] Robertson, S.: Integration von menschlicher und künstlicher Intelligenz zur Entscheidungsfindung im Leistungssport. In: Memmert, D. (eds) Künstliche Intelligenz und maschinelles Lernen in der Sportwissenschaft. Springer Spektrum, Berlin, Heidelberg, 2025. https://doi.org/10.1007/978-3-662-68950-9_6

[12] Nagel, L.; Lycklama, D.: How to Build, Run, and Govern Data Spaces. In: Otto, B., ten Hompel, M., Wrobel, S. (eds) Designing Data Spaces . Springer, Cham, 2022. https://doi.org/10.1007/978-3-030-93975-5_2

[13] Gersch, M.; Kari, A.; Schurig, T.; Lauf, F.; Wagener, H.: Europäischer Gesundheitsdatenraum: die zukünftige Ausgestaltung digitaler Infrastrukturen. In: Gemeinschaftsprojekt Gesundheit: Wie Value-Based Care das Gesundheitswesen neu erfindet (Hrsg: Hager, L), Berlin: MWV Medizinisch Wissenschaftliche Verlagsgesellschaft, 2024, S. 235–244.

[14] sphin-X e.V. https://sphin-x.de/ [abgerufen 28.01.2026]

DOI 10.5162/iCCC2026/4.3

Untersuchung der optischen Eigenschaften von Palladiumdünnschichten in Wasserstoff für einen Chip-basierten MOEMS-Wasserstoff-Sensor

Characterizing Optical properties of Palladium thin films in Hydrogen for chip-based MOEMS hydrogen sensor

Phil Frenzel, Robert Täschner, University of Applied Sciences Zwickau, 08056 Zwickau, Germany
phil.frenzel@whz.de

Kurzfassung

Um den Bedarf an Wasserstoff-Sensorik für sauerstofffreie Atmosphären (Elektrolyseure, Gasverteilnetz) zu decken, wird ein innovativer Chip-basierter MOEMS-Wasserstoff-Sensor entwickelt, welcher die Änderung der optischen Transmission von Palladium unter variierenden Wasserstoff-Konzentrationen ausnutzt. Als Basis für den Einsatz der Dünnschichten in der sensorischen Anwendung wurden die optischen Eigenschaften von Palladium bestimmt und werden in diesem Beitrag diskutiert. Weiterführend sollen diese Erkenntnisse in den Kontext der Anwendung gesetzt und erste Annahmen zu Eigenschaften des Sensors getroffen werden.

Abstract

To overcome the shortage of hydrogen sensors for oxygen-free environments (electrolyzers, pipelines, etc.), we develop an innovative chip-based MOEMS hydrogen sensor using the change in the optical transmission of palladium at varying hydrogen concentrations. As a first milestone, the optical properties of palladium thin films have been characterized and will be discussed in this paper. Further, the findings are set in context to the sensor system for first assumptions on the expected sensor performance.

1 Introduction

The Palladium-Hydrogen system is a thoroughly studied system with distinctive modifications, that reveals reversible volumetric, electrical and optical changes for varying hydrogen concentrations [1]-[3]. These phenomena propagate the use for sensor systems to measure hydrogen fractions in gas mixtures. Among others, the first concept to exploit the changes in optical transmission was given by [4]. In the scope of optical sensors, a vast variety of systems were developed. Those systems are using changes in transmission or other optical phenomena and the structure of the palladium-based active layer in different compositions, with or without optical fibers [5][6].
In our setup, a palladium layer on a photodiode is used as a switchable mirror. The photocurrent correlates to the irradiation intensity, that is determined by the palladium layer and varies upon hydrogen interaction as a result. The irradiation source is a bare-die LED placed and bonded on top of the final chip with a radiant power of 6.5 mW. To compensate for varying irradiation intensity, a second photodiode is intended as a reference, a third photodiode is blocked with an aluminum layer for offset correction, and a fourth photodiode is used optionally for an additional measurement layer or layer stack. This setup could be referred to as a four-quadrant photodiode with a hybrid-packaged light source (see Fig. 1). In contrast to previous developments, our system aims to be a wafer-based, scalable and low-cost option suitable for oxygen-free environments with the potential to measure arbitrary hydrogen concentrations also above the LEL [7].

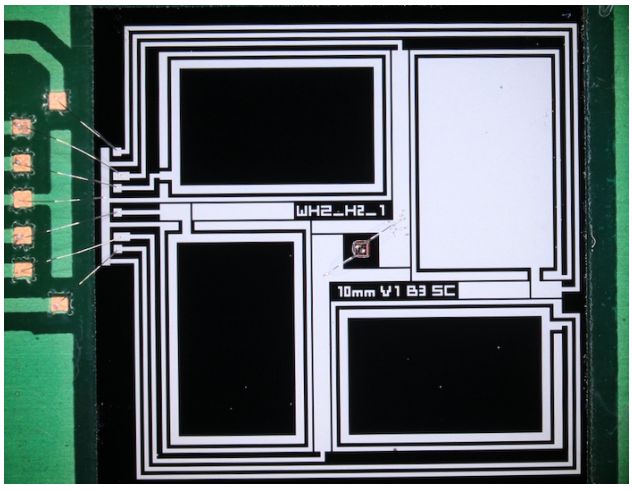


Fig. 1 Final sensor chip with four photodiodes and LED, mounted and bonded, chip size is approx. 12.6 x 12.6 mm^2

This paper reports on the basic examination of the transmission of palladium for various layer thicknesses and hydrogen concentrations. Notable observations along those measurements, concerning the formation of cracks in the film and basic dynamic properties, will be mentioned in brief. For the sensor design, the goal of this paper is to compare these findings to the spectral sensitivity of the silicon-based photodiode to derive a suitable wavelength for the

DOI 10.5162/iCCC2026/4.3

LED on the MOEMS-chip that promises the maximum responsivity for the complete system.

2 Experimental Procedure

To assess their optical transmission, the palladium layers were deposited on samples of float glass and fused silica (18 x 18 x 0.5 mm^3) by RF-Sputtering (Bias 1000 V / 0,25 Pa / 25 sccm Ar) from a 3-inch palladium-target (99.95 % w/w, FHR Anlagenbau GmbH, Ottendorf-Okrilla). The thickness of the film was determined by extrapolating the sputter time scaled according to thicker samples measured by a surface profilometer (Dektak 150, Veeco).
The coated samples were placed inside a custom build, sealed aluminum chamber (approx. 150 cm^3) with two float glass windows (Edmund Optics BV, Best, The Netherlands). To enable the evaluation of changes in transmission, the chamber can be evacuated and flooded from a gas bag reservoir under atmospheric pressure or constantly purged with a gas mixture at varying flowrates. The gas mixture is provided by a custom build setup, where two MFCs are controlled analogously to mix nitrogen and premixed Argon-Hydrogen gas (0.95 / 0.05). A correction for the nitrogen-calibrated MFCs was applied.
This setup is placed inside a Spectrophotometer JASCO V-770 (JASCO Deutschland GmbH, Pfungstadt) to measure the transmittivity in a dual beam configuration, meaning a reference beam is tracked during the measurement to compensate for changes in the light source or the detector [8]. A correction for the transmission of the substrate in the chamber is calculated afterwards to account for losses regarding the chamber windows and the substrate and a smoothing is applied for readability. However, the fraction of light that is reflected on the first surface (from air to substrate) is not considered for being absent in the case of coated substrates. This results in a systematic error. A random error occurs from the manual installation of the setup because of geometric offsets.

3 Results

Fig. 2 shows the transmission of the Pd layers vs the wavelength in nm for thicknesses of 15 nm to 60 nm in nitrogen and in a mixture of argon and hydrogen (0.95 / 0.05). The same measurements are displayed in Fig. 3 as relative change in transmission (ratio of transmission with 5 % v/v hydrogen to without hydrogen). Transmission decreases exponentially with increasing thickness following Lambert-Beer law. The relative change in transmission increases for increasing palladium height.
It should be pointed out, that layers higher than 30 nm result in macroscopic cracks at the first exposure to hydrogen (5 % v/v in Argon), significantly affecting the repeatability. For these layers, the measurements also show a decreased signal-to-noise ratio according to the lower transmission.

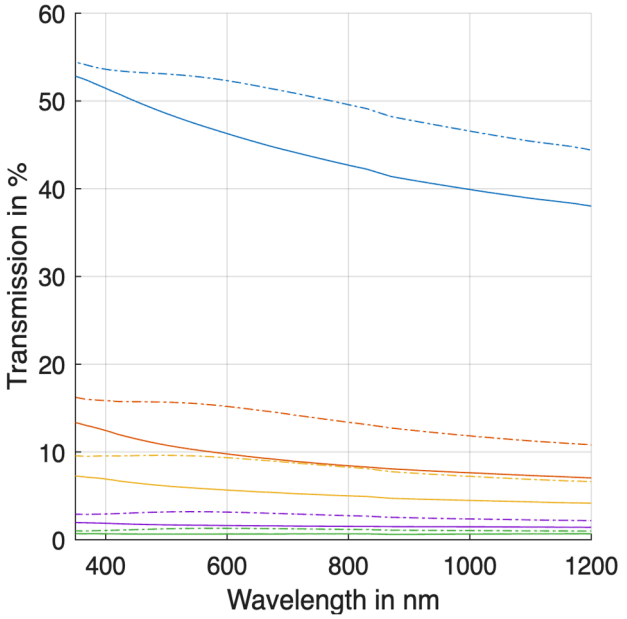


Fig. 2 Transmission of Palladium on Float Glass at 0 % (solid line) and 5 % v/v hydrogen in Ar (dash-dotted line) over the wavelength in nm for various thicknesses of Pd: 15 nm (blue), 22 nm (orange), 30 nm (yellow), 45 nm (purple), 60 nm (green).

For the relative change in transmission of layers thicker than 15 nm, a peak can be found below 800 nm. As the thickness increases, this peak develops more and shifts towards shorter wavelengths.

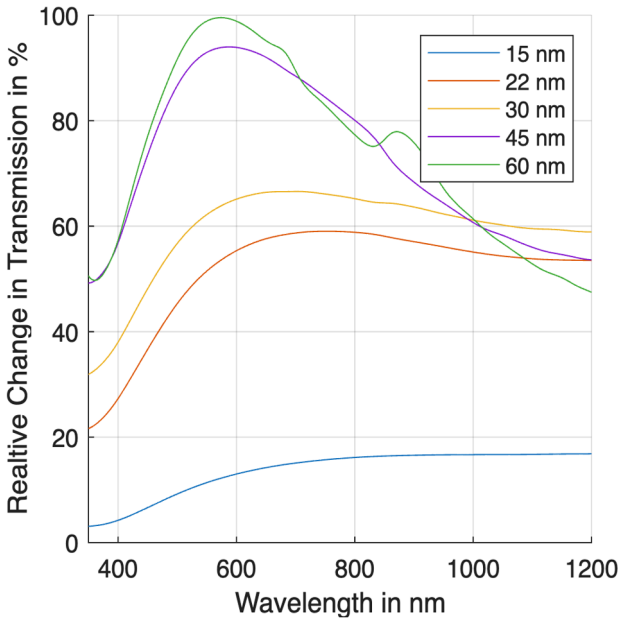


Fig. 3 Relative Change in transmission of Palladium on Float Glass for various thicknesses at 5 % v/v hydrogen in Ar compared to 0% v/v hydrogen over the wavelength in nm

To select an optimal wavelength for the 15 nm palladium layer, the spectral sensitivity of a typical silicon photodiode is plotted in comparison in Fig. 4. Regarding a limited selection of available LEDs, a wavelength of 850 nm was selected for a top-contact bare-die LED (Chips 4 Light

GmbH, Sinzing). For comparison, the absolute transmission of the 15 nm palladium layer is plotted against varying hydrogen concentrations from 0 to 5 % v/v in Fig. 5. For this layer, the relative change in transmission is more than 10 % compared to a hydrogen absent atmosphere.

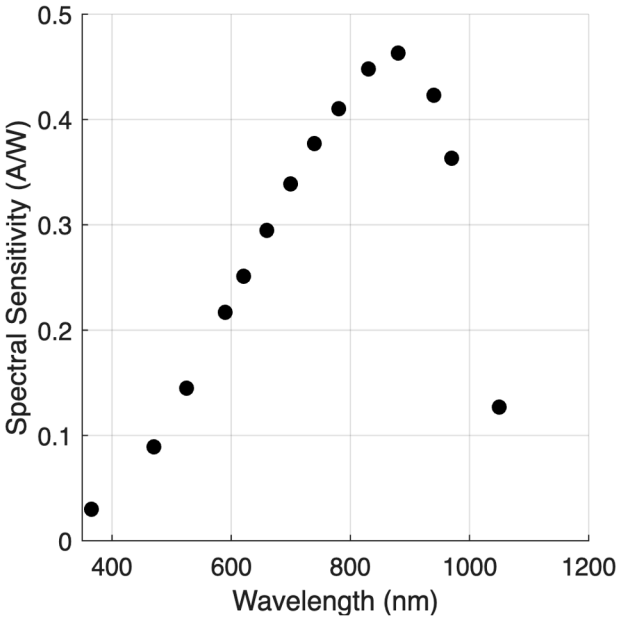


Fig. 4 Spectral Sensitivity in A/W over the wavelength in nm for a silicon photodiode (area 20 x 20 mm^2) as manufactured, referencing by measurements against calibrated equipment

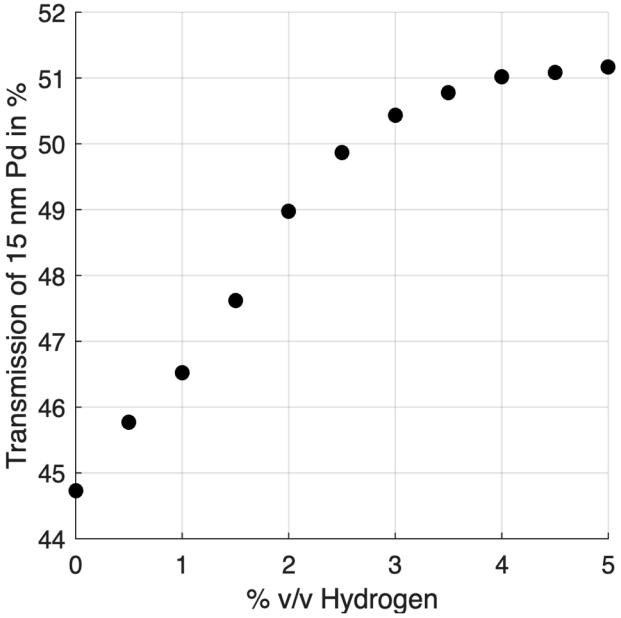


Fig. 5 Optical transmission at 850 nm of a 15 nm Pd layer on Float Glass against varying hydrogen concentrations, measured while purging at 1 slm.

During these examinations, some samples were stored in the same hydrogen containing atmosphere for 12 h. For that conditioning, also the 15 nm films show macroscopic cracks. However, these defects did not affect the change in transmission significantly. At a hydrogen concentration of 5 %, the deviation in the relative change were in range of +/- 3 % for the conditioned samples compared to the samples as manufactured. However, we noticed a change in the response time of the transmission (see Fig. 6), which will be discussed in the following.

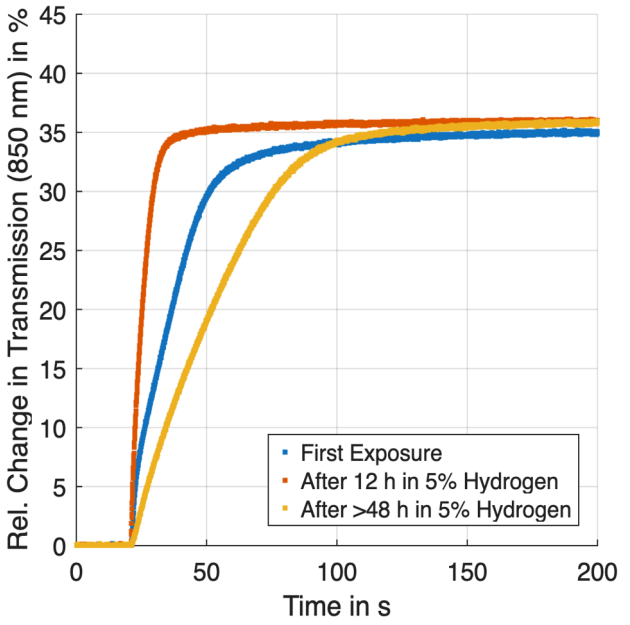


Fig. 6 Relative Change in transmission at 850 nm at a 5 % v/v hydrogen concentration compared to the transmission at 0 % v/v hydrogen transmission over time. The chamber was evacuated and afterwards flooded at 20 s with the gas mixture to atmospheric pressure.

4 Discussion

A change in transmission over the concentration of hydrogen (more precise: its partial pressure) as depicted in Fig. 5 can be related to a change in parameters of the microlattice of Pd [1]. These modifications contribute to the optical and mechanical phenomena observed in the assessed concentration range. The change in transmission flattens for concentrations higher than 3.5 %. This matches previous findings, although some change should be detectable up to 100 % with pure Pd [9]. If these changes are not sufficient for the use in the actual measurement setup, an active layer either with material compositions or layers of other materials with a Pd capping layer can be used as implemented in [10].

Regarding the mechanical stability and crack formation of the films, the effects on the sensor performance should be questioned. If a continuous exposition to hydrogen atmospheres does not affect the sensor performance, further steps may be oblique. This needs to be tested in following studies, e.g. at various concentrations and temperatures. To compensate these flaws, an adhesion layer of titanium underneath the Pd layer could be used, although its thickness should be optimized in respect to the optical transmittance. Another option might be the application of a nanostructure, which can not only improve adhesion, but could also enhance the dynamic properties of the film's transmission change.

Up to now, no evaluation of rise and fall times were successful and reproducible despite different measurement settings (flooding or purging) probably due to thermal effects during irradiation. Following established procedures, the transmission was measured against time for varying mixtures of gas while purging the chamber. Since a systematic error occurs from the time to establish the intended concentration inside the chamber, another method was used. Instead, the chamber was evacuated first with a diaphragm pump and then flooded with a gas mixture from a gas bag reservoir at atmospheric pressure as stated above. Even for this procedure, no correlation was found between the varying response times of the transmission change and other parameters. Recent additional measurements point towards the influence of the Pd film's temperature, since the irradiation of the samples took place during the evacuation. different substrate temperatures could be a consequence. If this is true, it could lead to the different response times and could explain the missing correlation [11]. This needs to be validated by additional measurements with a constant irradiation time.

The use of nanostructures in the sensor design could offer advantages in the film stability, higher absolute transmission and response times < 5 s as preliminary tests showed. To compensate for lower transmission when using multiple layers, the decrease in reflectivity by sub-wavelength structures can be of interest. One promising approach is the inhouse full wafer pattering via Displacement Talbot Lithography that is currently researched [12].

5 Conclusion and Outlook towards sensor performance

The acquired transmission data supports the intended working principle of the sensor in oxygen-free environments in the range of up to 5 % v/v hydrogen. A non-linear response can be assumed as expected. The mechanical stability of the palladium layers should be evaluated for the sensor system regarding a required preliminary conditioning of the films.

Besides the mentioned considerations about the layer, the electronic sensor system will be completed in the distant future. This includes, inter alia, the design and manufacturing of a PCB based on an ATmeag328P with an appropriate Transimpedance-Amplifier of suitable gain, an ADC device as well as an I^2C-Interface.

For the final system, the response time will be evaluated at higher hydrogen concentrations and/or with different background gases (possibly including CO, CO_2, H_2S, CH_4, H_2O vapor, among others). Finally, a temperature-dependent calibration is required and an option to measure the temperature on the sensor by resistivity or forward bias shift should be considered.

6 Acknowledgements

This research was co-funded by the European Union and by tax revenue based on the budget of the Free State of Saxony under ESF Plus Grant No. 100693823. The optical measurements took place at Fraunhofer AZOM, Zwickau, with help from an astonishing team. This opportunity is gratefully acknowledged. The various help from all members of the department of MEMS and Nanolab at UAS Zwickau is appreciated.

7 Literature

[1] Manchester, F. D et al.: The H-Pd (Hydrogen-Palladium) System. Journal of Phase Equilibria 15 (1), 1994, p. 62–83.

[2] Hübert, T. et al.: Hydrogen Sensors – A Review. Sensors and Actuators B: Chemical 157 (2), 2011, p. 329–52.

[3] Palm, K. J. et al.: Dynamic Optical Properties of Metal Hydrides. ACS Photonics 5 (11), 2018, p. 4677–86.

[4] Butler, M. A.: Micromirror Optical-Fiber Hydrogen Sensor. Sensors and Actuators B: Chemical 22 (2), 1994, p. 155–63.

[5] Darmadi, I. et al.: Bulk-Processed Plasmonic Plastic Nanocomposite Materials for Optical Hydrogen Detection. Accounts of Chemical Research 56 (13), 2023, p. 1850–61.

[6] Bramann, G. et al.: New Optical Hydrogen Sensor with Fast Response Time Based on Multilayer Palladium-Nickel-PTFE Thin Film for Explosion-Proof Detection of High H2 Concentrations of 1-100%. Third European Workshop on Optical Fibre Sensors, Juli 4th, 2007, 66191L.

[7] Schröder, V. et al.: Sicherheitstechnische Eigenschaften von Erdgas-Wasserstoff-Gemischen. Abschlussbericht zum Forschungsvorhaben 2539, Bundesanstalt für Materialforschung und –prüfung (BAM). Berlin, 2016.

[8] https://jascoinc.com/products/spectroscopy/uv-visible-nir-spectrophotometers/models/v-770-uv-visible-nir-spectrophotometer/. Accessed on: 24.11.2025.

[9] Mandelis, A. and Garcia, J. A.: Pd/PVDF Thin Film Hydrogen Sensor Based on Laser-Amplitude-Modulated Optical-Transmittance: Dependence on H2 Concentration and Device Physics. Sensors and Actuators B: Chemical 49 (3), 1998, p. 258–67.

[10] Bannenberg, L. et al.: Tantalum-Palladium: Hysteresis-Free Optical Hydrogen Sensor Over 7 Orders of Magnitude in Pressure with Sub-Second Response. Advanced Functional Materials 31 (16), 2021, p. 2010483.

[11] Shutaev, V. A. et al.: Kinetics of Changing in Optical Transmittance of Palladium Nanolayers during Interaction with Hydrogen. Optics and Spectroscopy 131 (3), 2023, p. 178–82.

[12] Solak, H. H. et al.: Displacement Talbot lithography: a new method for high-resolution patterning of large areas. Opt. Express 19, 2011, p. 10686-10691.

DOI 10.5162/iCCC2026/4.4

Voll digitalisiertes Reallabor Wasserstofftankstelle der BAM

Robert Bock, Bundesanstalt für Materialforschung und -prüfung (BAM), Berlin, Deutschland
robert.bock@bam.de;
Carlo Tiebe, Bundesanstalt für Materialforschung und -prüfung (BAM), Berlin, Deutschland
carlo.tiebe@bam.de
Finn Lückoff, Bundesanstalt für Materialforschung und -prüfung (BAM), Berlin, Deutschland
finn.lueckoff@bam.de
Georg Mair, Bundesanstalt für Materialforschung und -prüfung (BAM), Berlin, Deutschland
georg.mair@bam.de

Kurzfassung

Moderne digitale Technologien bieten neue Möglichkeiten für das Sicherheitsmanagement und die Qualitätssicherung von technischen Anlagen. Aus der Initiative QI-Digital hervorgehend zeigt die BAM im Reallabor Wasserstofftankstelle, wie digital gestützte Verfahren und Werkzeuge der Qualitätsinfrastruktur (QI) zu einer effizienteren und transparenteren Qualitätssicherung von technischen Anlagen beitragen. Damit kann die Anlage wirtschaftlicher und verlässlicher betrieben werden. Das Reallabor bildet die gesamte Wasserstoffwertschöpfungskette ab – mit der Anlage wird Wasserstoff erzeugt, gespeichert und an verschiedene Fahrzeugtypen abgegeben.

Abstract

Modern digital technologies offer new opportunities for safety management and quality assurance in technical facilities. As part of the QI-Digital initiative, BAM is demonstrating in its living lab hydrogen refueling station how digitally supported quality infrastructure (QI) processes and tools can contribute to more efficient and transparent quality assurance in technical facilities. This allows such facilities to be operated more economically and reliably. The living lab maps the entire hydrogen value chain – the plant produces hydrogen, stores it, and dispenses it to various types of vehicles.

1 Motivation

Herkömmliche Inspektions- und Wartungsverfahren komplexer technischer Anlagen sind oft aufwändig und führen zu langen oder kurzfristigen Stillstandszeiten. Es ist derzeit kaum möglich, den Zustand der Gesamtanlage oder einzelner Komponenten sowie den Kalibrierstatus der Sensoren umfassend und in Echtzeit zu erfassen und zu dokumentieren. Häufig existieren lediglich asynchrone und oft analoge Sammlungen von qualitäts- und sicherheitsrelevanten Dokumentationen, die umfangreich und unübersichtlich sind. Dies sind nur einige der typischen Herausforderungen der Qualitätssicherung technischer Anlagen.
Am Beispiel einer Wasserstofftankstelle entwickelt, erprobt und demonstriert die BAM Lösungen für eine moderne QI, die den gestiegenen Anforderungen des Betriebs komplexer technischer Anlagen gerecht wird. Neben neuen Überwachungs- und Prüfmethoden, Sensortechnologien und digitalen Abbildern, geht es dabei insbesondere um die Entwicklung und Erprobung von grundlegenden Werkzeugen einer digitalen QI, wie maschinenlesbaren Standards, digitalen Prüfberichten und Zertifikaten, und einer übergreifenden IT- und Datenplattform [1].
Die Lösungen und Entwicklungen für eine digitale QI können zur Optimierung von Qualitätssicherungsprozessen an Wasserstofftankstellen beitragen. So kann durch ihre Implementierung der Betrieb zuverlässiger, sicherer und wirtschaftlicher gestaltet und Ressourcen effizienter eingesetzt werden. Dies wirkt sich positiv auf die Akzeptanz von Wasserstoffanwendungen, insbesondere die wasserstoffbasierte Mobilität, aus. Darüber hinaus sind die QI-Werkzeuge, Prozesse und Verfahren so entwickelt, dass diese grundsätzlich übertragbar sind - auf weitere Wasserstoffanwendungen und technische Anlagen allgemein. Um eine umfassende digitale QI zu ermöglichen, werden die im Reallabor gewonnenen Erkenntnisse in die Weiterentwicklung von technischen Normungs- und Rechtsvorschriften eingebracht.

1.1 Die digitale Abbildung des Gesamtsystems

Ein grundlegender Schritt hin zu einem integrierten Überwachungskonzept technischer Anlagen ist deren digitale Abbildung. Dafür werden Anlagenzustand und Umgebungsbedingungen mithilfe der eingesetzten Sensoren und Aktuatoren charakterisiert und entsprechend implementiert. Kern dessen ist eine moderne Prozessleittechnik (PLT). Sie erfasst die aktuellen Überwachungsdaten aus den Sensornetzwerken und interpretiert, verarbeitet und bewertet diese in einem digitalen Zwilling in Echtzeit für ein modernes Zustandsmanagement und eine vorausschauende Wartung der technischen Anlage.
Im Reallabor Wasserstofftankstelle (Abbildung 1) entwickeln und erproben wir ein PLT- und Sensorsystem prototypisch, das in der Lage ist, alle qualitätsrelevanten Informationen und Nachweise (vom digitalen Konformitätsnachweis (D-CoC) der eingebauten Komponenten bis hin zu Kalibrierzertifikaten (DCC) verwendeter Messsysteme [2]) ins System einzubinden und deren Rückverfolgbarkeit zu gewährleisten. Dies wird auf Basis der sogenannten Verwaltungsschale erprobt, die das Rückgrat der Industrie 4.0 bildet und oft als Digitaler Zwilling bezeichnet wird. Die Verwaltungsschale ist ein branchenneutraler Standard,

DOI 10.5162/iCCC2026/4.4

der es ermöglicht, dass „Assets“ (Anlagen oder Anlagenteile) auf der ganzen Welt miteinander kommunizieren können.
Diese Informationen können dann relevanten Stakeholdern (z.B. Marktüberwachungs- und Konformitätsbewertungsstellen) auf kontrollierten Zugriffsebenen sicher, datensouverän und adressatenorientiert zur Verfügung gestellt werden. Die Verbindung von Softwarediensten mit den offenen, standardisierten Protokollen und Interaktionssequenzen der Datenräume (Data Spaces) zum Austausch QI-relevanter Assets wird unter dem Arbeitstitel „Quality-X“ behandelt.

1.2 Sensoren und Sensornetzwerke für die digitale QI

In technischen Anlagen sind in großer Zahl sicherheitstechnisch relevante Sensoren verschiedener Art und zunehmend komplexe Sensornetzwerke installiert. Im Reallabor werden Lösungen entwickelt und erprobt, die deren verlässlichen Betrieb und effektives Management mit Hilfe von digitalen QI-Werkzeugen und – Prozessen ermöglichen.
Durch den Einsatz von digitalen Kalibrierscheinen (DCC) und Konformitätsnachweisen (D-CoC) kann die bisher manuelle Integration in die Prozessleittechnik (PLT) in Zukunft digital automatisiert geleistet werden. Dies erleichtert den Aufbau komplexer Sensornetzwerke sowie den automatischen Vergleich von Messwerten und eine kontinuierliche Plausibilitätsprüfung. Der Ausfall einzelner Sensoren kann durch die hohe Informationsdichte eines Sensornetzes kompensiert werden. Eine hohe Sensordichte ermöglicht außerdem Wasserstoffverluste und Fehlfunktionen frühzeitig zu erkennen und zu orten. Dies wird im Reallabor dafür genutzt, ein operationscharakteristisches Leckagedetektions-/Explosionsschutzkonzept zu entwickeln und erproben.
Darüber hinaus entwickeln und erproben wir sensorgestützte Lösungen für die digitale Überwachung der Reinheit des Wasserstoffs auf dem Weg durch die Tankstelle. Entsprechende Qualitätsnachweise sind nicht nur für Kunden bedeutsam, sondern dienen auch dazu Schäden frühzeitig zu erkennen und Verunreinigung des Gases zu vermeiden, durch z.B. innere Undichtigkeiten, Abrieb oder Schmiermittel. Somit können Systemkomponenten zuverlässiger betrieben und einem frühzeitigen Systemausfall entgegengewirkt werden [3].

1.3 Digitale Überwachung von Anlagenkomponenten am Beispiel der H2-Druckbehälter

Eine wesentliche Komponente von Industrieanlagen sind Druckbehälter. Momentan werden solche Druckgasspeicher in stationären Anlagen periodisch geprüft, ohne die Nutzungshistorie zu beachten. Ziel der digitalen QI ist hingegen eine kontinuierliche und betriebsbegleitende Überwachung der Bauteilintegrität. Damit können die Prüf- und Kalibrierfristen von reinen Zeitintervallen in Fristen umgewandelt werden, die von der Belastung und dem Zustand abhängig sind.
Dazu werden neue, digital-gestützte Überwachungsmethoden unter Nutzung von digitalen Zertifikaten entwickelt und erprobt. Auf Basis leistungsstarker zerstörungsfreier Prüfmethoden kann in Verbindung mit einer (künstlich) intelligenten Auswertung der aktuelle Zustand kontinuierlich und umfassend bewertet werden. Ein solches KI-unterstütztes Structural Health Monitoring (SHM) erlaubt dann eine laufende Restlebensdauerabschätzung für den relevanten Druckbehälter, im Idealfall sogar nach probabilistischen Aspekten. So kann die Betriebssicherheit und ggf. die Restlebensdauer laufend beurteilt werden. Dafür werden neben den technischen Aspekten auch notwendige Anpassungen von Normung und Regelsetzung untersucht.

2 Anlagenbeschreibung

Das Reallabor [4] besteht aus Komponenten zur Wasserstofferzeugung mittels Elektrolyse bzw. Wasserstoffbefüllung über Trailerversorgung. Darüber hinaus werden Mitteldruckspeicher, Verdichter, Hochdruckspeicher, Wärmeübertrager sowie eine Dispensereinheit zum Befüllen von Fahrzeugen über H70 und H35 Leitungen eingesetzt (siehe Abbildung 2).
Der in der Anlage eingesetzte Wasserstoff kann mittels einer eigenen Elektrolyseeinheit (bis zu 24 kg_{H2}/Tag) vor Ort unter Einsatz einer lokalen PV-Anlage erzeugt und direkt bei Mitteldruck (200 bar) gespeichert werden. Alternativ kann dieser Speicher auch von einem Wasserstofftrailer versorgt werden. Mittels Kompressors wird der Wasserstoff im Hochdruckspeicher auf 930 bar weiter verdichtet. Die nachgeschaltete Kühleinheit konditioniert den Wasserstoff derart, dass Fahrzeuge bei 350 und 700 bar über den Dispenser betankt werden können. Das Reallabor ist in modularer Containerbauweise realisiert worden.

Abbildung 1: Betankung des BAM-Wasserstofffahrzeugs im gegenwärtigen Aufbau des Reallabors

DOI 10.5162/iCCC2026/4.4

Abbildung 2: Hauptkomponenten und Wertschöpfungskette des Reallabors Wasserstofftankstelle.

Die Prozessüberwachung, Dispenser, Elektrolyse, Speicher und Kompressor sind in separaten Containern untergebracht und können damit bedarfsgerecht auf dem Testgelände Technische Sicherheit (TTS) der BAM aufgestellt werden (siehe Abbildung 3).
Der Aufbau und die Komponentenauswahl des Reallabors sind derart gewählt, dass projektspezifische Erweiterungen bzw. Umbauten möglich sind. Damit ergeben sich besondere Ausstattungsmerkmale:

- (H_2-)Sensornetzwerk [5]
- Prozessüberwachung nach Industrie 4.0 [6]
- IT/OT-Infrastruktur für QI-Digital Werkzeuge [6]
- Digitaler Zwilling mit Prozessmodellierung
- Großflächige Aufstellung, baulicher Schutz
- Autarke und grüne Wasserstofferzeugung
- Gas-Probenahmestellen
- Einbindung weiterer Komponenten möglich

3 Datenmanagement

Die digitale Qualitätsinfrastruktur unterliegt einer umfassenden Datenintegration, Auswertung und Lesbarkeit im Hinblick auf Interoperabilität. Die dafür erforderliche IT/OT-Infrastruktur bildet das Rückgrat für die durchgängige Integration von Sensordaten – von der Feldebene, auf der zuverlässige und echtzeitfähige Operational Technology (OT) eingesetzt wird, bis zur Anwendungsebene (IT), auf der die Daten verwaltet, überwacht und für die Weiterverarbeitung bereitgestellt werden [6].
Die IT-Technologie ist verantwortlich für das Datenmanagement, Anwendungen zur Bewertung der Qualitätsinfrastruktur sowie die Cybersicherheit. Darüber hinaus stellt die IT die Netzwerksicherheit und den Schutz der Komponenten sicher. Die Operational Technology (OT) hingegen ist verantwortlich für die Steuerung der Sensoren und Aktoren im Reallabor sowie für die Betriebssicherheit.

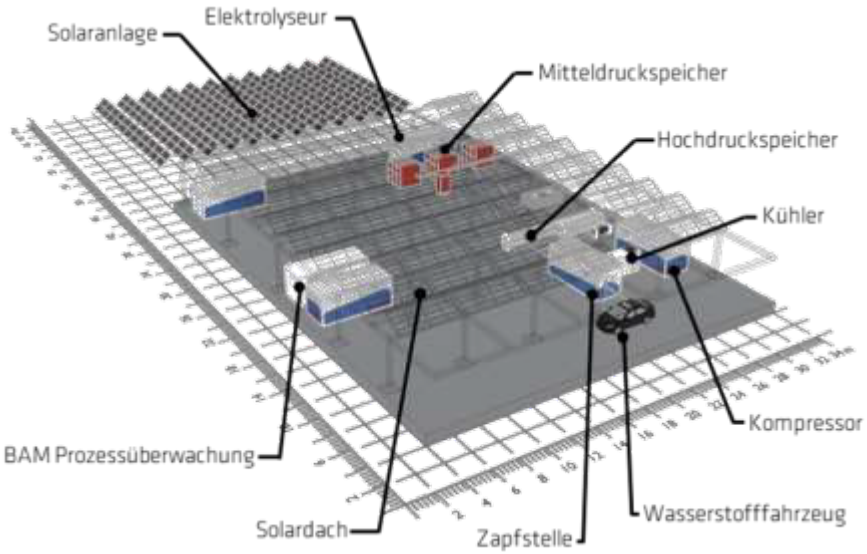


Abbildung 3: Aufbau des Reallabors mit dessen Hauptkomponenten

DOI 10.5162/iCCC2026/4.4

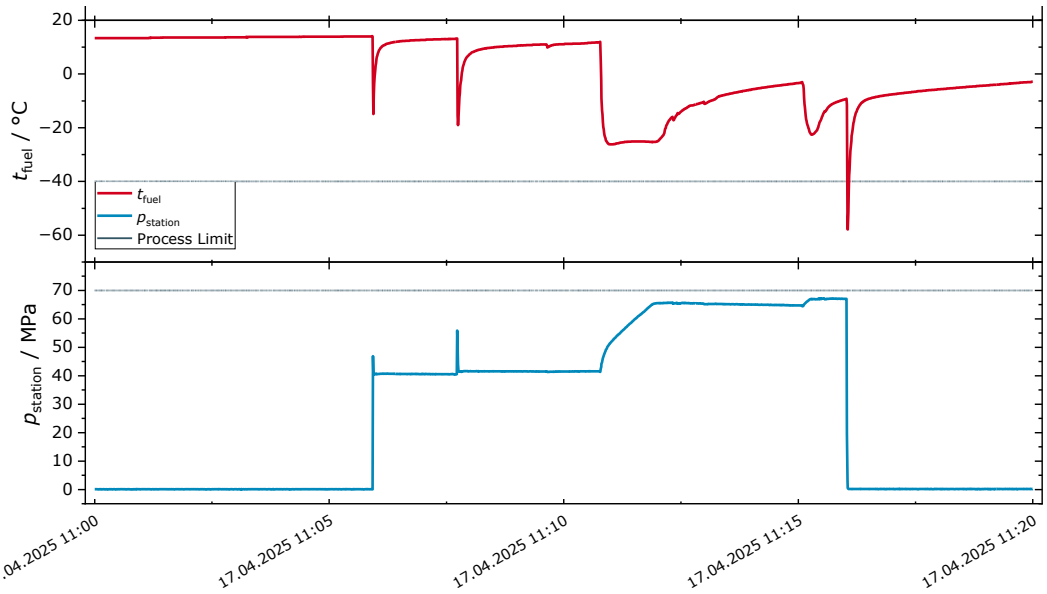


Abbildung 4: Sensordaten eines Betankungsvorgangs, der die gesetzten Prozessgrenzen überschritten hat und abgebrochen wurde. In dieser Darstellung sind die Wasserstoffabgabetemperatur (t_{fuel} / °C) sowie der Gasdruck am Dispenser ($p_{station}$ / MPa) aufgetragen.

Interoperabilität gemäß den Anforderungen von Industrie 4.0 wird durch ein standardisiertes Datenmanagement mit einheitlicher Struktur und Semantik gewährleistet.
Durch die Kombination des IT/OT-Backbones mit den oben genannten QI-Digital-Werkzeugen (DCC, D-CoC, VWS) wird eine Echtzeitüberwachung gemäß Industrie 4.0 Standard ermöglicht. Überschreitungen der Prozessgrenzen werden nach industriellem Standard auch im Reallabor schnell erkannt (Abbildung 4), und die Daten darüber hinaus bei Bedarf auch Stakeholdern in Echtzeit zur Verfügung gestellt. Alle relevanten Assets des Reallabors können somit nach diesem Standard abgebildet werden. Mittels REST-API und browserbasiertem Portal ist ein Fernzugriff auf alle Sensordaten möglich. Daraus ergeben sich folgende Besonderheiten der IT/OT:

- Standortübergreifende Konnektivität
- Funktionale Sicherheit des Systems
- Cybersicherheit
- Datenverfügbarkeit und -verfolgbarkeit
- Standardisierter Digitaler Zwilling (Verwaltungsschale VWS)

Die Dateninfrastruktur des Reallabors ermöglicht es einen Digitalen Zwilling der Anlage zu erzeugen. Dieser stellt in einer 3D-Umgebung einen digitalen Zugang zur Gesamtanlage bereit (Abbildung 5). Basierend auf umfangreichen Datensätzen aus dem spezifischen IT/OT-Backbone können somit weitreichende Prozesssimulationen durchgeführt werden.

Abbildung 5: Digitaler Zwilling mit 3D-Umgebung und Prozesssimulation

4 Vision

In Anbetracht der vor uns liegenden Energiewende mit umfangreichen und komplexen Anforderungen an die verschiedenen Energieträger trägt das Reallabor dazu bei, den zukünftigen Energiemarkt zu gestalten. Um dem gerecht zu werden, liegt es nahe, über die Erzeugung, Speicherung und Nutzung von Wasserstoff hinauszugehen und andere Energiespeicherarten (v.a. Batterien) mit einzubeziehen. Die damit entstehende Sektorenkopplung im Zusammenspiel mit der digitalen Infrastruktur (IT/OT, Digitaler Zwilling etc.) bietet ein breites Spektrum an Forschungs- und Anwendungsmöglichkeiten. Gemeinsam mit externen Forschungspartnern, der Industrie aber auch KMUs und Start-Ups wollen wir das einzigartige Reallabor weiterentwickeln und nutzen. Damit leisten wir einen entscheidenden Beitrag zu einer sicheren Energieversorgung der Zukunft.

5 Literatur

[1] Initiative QI-Digital, Pilotprojekt "Verlässliche Wasserstofftankstelle" - eine digitale Qualitätsinfrastruktur (QI) für technische Anlagen *Website* **2025–12–01 (Aufrufdatum)**, https://www.qi-digital.de/pilotprojekte/verlaessliche-wasserstofftechnologie.

[2] N. Brunner, M. Melzer, Digital Certificates: Enabling Automation in Quality Assurance and Metrological Traceability *Sensor Measurement Science International (SMSI 2025)* **2025**, 171–172, https://doi.org/10.5162/SMSI2025/C6.4.

[3] C. Schaad, C. Tiebe, K. M. Groth, A Framework for Transforming Process Control System Data from a Hydrogen Fueling Station into HyCReD Data *Proceedings 35th European Safety and Reliability Conference (ESREL2025) and the 33rd Society for Risk Analysis Europe Conference (SRA-E 2025)* **15.–19.06.2025**, *ESREL 2025, Stavanger, Norway*, https://doi.org/10.3850/978-981-94-3281-3-procd.

[4] Bundesanstalt für Materialforschung und -prüfung (BAM), Digitale Qualitätsinfrastruktur für Wasserstoff: BAM eröffnet einzigartige Forschungstankstelle *Pressemitteilung* **2025–07–03**, https://www.bam.de/Content/DE/Pressemitteilungen/2025/Energie/2025-07-03-wasserstofftankstelle.html.

[5] K. Nattuveettil, C. Liebner, C. Tiebe, Hydrogen safety- Dealing with Closed Spaces *SMSI 2025* **2025**, 167–168, https://doi.org/10.5162/SMSI2025/C6.2.

[6] M. Maiwald, R. Soruco Aloisio, K. Meyer, C. Klaus, Auf dem Weg zu einer digitalen Qualitätsinfrastruktur - Eine Labor-Testplattform für die Integration von Sensoren und Messgeräten *GIT Labor-Fachzeitschrift* **2023**, *67*, 9, 20–23, https://analyticalscience.wiley.com/content/magazine-do/git-labor-fachzeitschrift-09-2023.

DOI 10.5162/iCCC2026/4.5

Portable NIR Spectroscopy for Reliable Characterization of High-Intensity Radiating Objects: Statistical vs AI-Based Modelling

Maedeh Nobari 1), 2), Friedrich Eckert, 2), Heinrich Engelke, 1), and Ireneusz Jabłoński 1), 2)
1) Integrated Silicon Systems, Fraunhofer Institute for Photonic Microsystems, Cottbus/Germany, maedeh.nobari@ipms.fraunhofer.de 2) Chair of Micro and Nano Systems, Faculty MINT, Brandenburg University of Technology Cottbus-Senftenberg, Germany

Abstract

Portable near-infrared (NIR) spectroscopy offers a powerful non-contact approach for characterizing high-temperature radiating surfaces, enabling in-situ temperature diagnostics without intrusive sensors. However, achieving measurement reliability under intense thermal emission is challenged by detector nonlinearity, emissivity variation, and calibration instability. This study introduces a portable NIR framework integrating a blackbody calibration device (BBCD) with advanced data-driven modelling to improve accuracy and reproducibility across the 600–1200℃ range. We systematically compare statistical regression methods with AI-based architectures-including fully connected, convolutional, residual, and transformer networks for temperature prediction from spectral data. Results show that AI models, particularly transformer and ResNet architectures, outperform statistical baselines in both accuracy and generalization, especially under irregular sampling and noisy conditions. The findings highlight the importance of BBCD assisted calibration and robust preprocessing in achieving traceable, field-ready temperature measurements, advancing the application of portable NIR spectroscopy in industrial and scientific diagnostics.

1 Introduction

Portable near-infrared (NIR) spectroscopy is an increasingly vital technology for the non-destructive, in-situ characterization of materials. Its adoption is rapidly growing across diverse fields, from agriculture and pharmaceuticals to industrial process control, driven by the development of robust, cost-effective handheld instruments. In industrial settings, this technology is particularly valuable for characterizing high-temperature, incandescent ("blackbody-like") objects, offering a safe and immediate method for monitoring processes reaching up to 1200°C. By analysing the thermal radiation emitted directly from a surface, portable NIR pyrometry eliminates the need for sample extraction and reduces operator exposure to hazardous environments, aligning with the principles of green analytical chemistry [1].

Despite its potential, achieving accurate and repeatable temperature measurements of incandescent emitters with portable NIR spectroscopy presents significant challenges. The measurement integrity is often compromised by factors such as instrumental drift, spectral noise, and fluctuating ambient conditions. More critically, the high-intensity radiation can lead to detector saturation, while the target's surface emissivity, its efficiency in emitting thermal energy can vary significantly with temperature and oxidation state, introducing a primary source of error [2].

To address these challenges in thermal analysis, blackbody calibration devices (BBCDs) serve as an essential reference standard. A BBCD provides a highly stable and uniform radiation source that approximates an ideal emitter (emissivity ≈ 1.0), allowing for the precise characterization and correction of the spectrometer's response. By calibrating the instrument against a traceable black-body source, it becomes possible to account for instrumental distortions and create a valid basis for converting spectral radiance measurements into accurate temperature predictions. This calibration step is critical for ensuring that the data collected in harsh field or factory environments is both reliable and comparable. By interleaving measurements of the unknown hot surface with calibration against BBCD spectra, one can better understand and constrain spectral system deviations and thus im-prove the robustness of downstream modelling.

Against this backdrop, the aim of the present study is to systematically compare two modelling paradigms—classical statistical calibration methods versus modern AI-based (machine learning or neural network) approaches—for the prediction of surface temperature from NIR spectral data of high-temperature radiating objects. We use a portable NIR spectrometer in conjunction with a calibrated BBCD reference to collect training and test spectra, then evaluate and benchmark the accuracy, repeatability, and robustness of both modelling approaches. Our goal is to assess which of these two approaches offers superior performance (or under what conditions each excels) in the context of high-intensity thermal emission, thereby guiding future design of portable NIR systems for high-temperature applications.

2 State of The Art

Portable NIR spectroscopy has rapidly evolved from compositional analysis to applications in thermal emission and temperature diagnostics. Advances in sensor miniaturization and calibration stability now enable reliable operation in high-temperature environments, supporting integration into industrial monitoring and quality control systems [3]. The fusion of NIR sensing with pyrometric modelling allows non-contact surface temperature estimation, providing a viable alternative to contact thermometry. Yet, most studies remain limited to laboratory-scale blackbody or bench-top systems, with few addressing the challenges of

true portability [4]. Transitioning to field-deployable instruments introduces new uncertainties that demand enhanced hardware precision and data-driven analytical approaches.

Despite recent advances, measurement accuracy and reproducibility in portable NIR pyrometry remain challenging. Variations in detector sensitivity, optical alignment, and temperature-dependent emissivity continue to limit precision [5]. Issues such as data scattering, non-linearity, and low signal-to-noise ratios further complicate regression-based temperature estimation. Moreover, calibration transfer between instruments is often unreliable, as environmental effects - vibration, ambient light, and thermal drift - distort spectral responses. Studies by Pohl et al. [5] and Yang et al. [6] show that even with well-characterized blackbody references, maintaining accuracy in field conditions demands dynamic calibration and uncertainty modelling. Hence, standardized calibration frameworks and adaptive, transfer-learning–based models are increasingly required to ensure robustness.

To overcome these limitations, recent research trends are converging toward the integration of blackbody calibration devices (BBCDs) with machine learning and deep learning models to enhance the reliability of thermal spectral analysis [7]. Blackbody-based calibration offers a traceable physical standard that constrains emissivity-related errors, while data-driven models can learn complex, nonlinear mappings between spectra and temperature more effectively than traditional regression [8]. Hybrid approaches that combine physical calibration and AI modelling—such as convolutional neural networks (CNNs) or hybrid chemometric–ML models—have demonstrated improved accuracy in infrared thermography and spectral pyrometry [9]. For instance, Wang et al. [7] used deep convolutional architectures to predict emissivity-corrected temperature fields from NIR spectra, outperforming polynomial regression and PLSR in both accuracy and generalization. Similarly, Zhang et al. [10] reported that transformer-based regression networks achieved higher robustness under varying optical noise conditions compared to statistical baselines. However, systematic benchmarking between purely statistical and AI-based modelling approaches under calibrated, portable measurement conditions remains limited—representing the critical research gap addressed in this study.

3 Materials and Methods

This study explores the integration of portable near-infrared (NIR) spectroscopy with advanced data-driven modelling to achieve reliable temperature estimation of high-temperature radiating (blackbody-like) objects. The work investigates how calibration accuracy, ensured through a Blackbody Calibration Device (BBCD), and model architecture selection influence prediction robustness. We compare the performance of statistical regression and AI-based neural models—including Fully Connected Networks (FCN), Convolutional Neural Networks (CNN1D), Residual Networks (ResNet1D), and Transformer-based architectures—across irregularly sampled temperature datasets (600–1200 °c).

3.1 Black Body Calibration Device and Measurement Reliability

The Blackbody Calibration Device (BBCD) serves as the primary radiometric reference for ensuring the accuracy and reliability of spectral temperature measurements. It provides a uniform and traceable radiation source approximating an ideal blackbody (emissivity 0.99 ± 0.01). The emitted spectral radiance is described by Planck's law:

$$L_\lambda(T) = \frac{2hc^2}{\lambda^5}\frac{1}{\exp\left(\frac{hc}{\lambda k_B T}\right)-1} \quad (1)$$

Where h is Planck's constant, c the speed of light, k_B the Boltzmann constant, λ the wavelength, and T the temperature.

Under high-radiance conditions (600-1200℃), portable NIR spectrometers can suffer from detector nonlinearity, optical scattering, and spectral drift. By calibrating each measurement against the BBCD spectrum, the system response function $H(\lambda)$ can be characterized:

$$H(\lambda) = \frac{I_{measurement}(\lambda)}{L_\lambda(T_{BBCD})} \quad (2)$$

Where $I_{measurement}(\lambda)$ is the raw detector signal. The corrected spectral intensity of an unknown emitter is then:

$$I_{corrected}(\lambda) = \frac{I_{sample}(\lambda)}{H(\lambda)} \quad (3)$$

This calibration minimizes systematic errors due to grating efficiency and detector sensitivity variations, ensuring that subsequent temperature predictions reflect physical surface emission rather than instrumental artifacts. The BBCD and setup it shown in Figure 1. For data acquisition, an SR-200N-33 black-body radiator was utilized, providing a temperature range of 50 to 1200 °C, with a thermal accuracy of ±2 °C and a thermal stability of ±0.50 °C. The aperture of the black-body radiator was equipped with a protective quartz glass window, a collimator (ThorLabs F260SMA-1550), and an optical fiber (ThorLabs M44L02), which was connected to the NIRQUEST NQ5200147 NIR spectrometer.

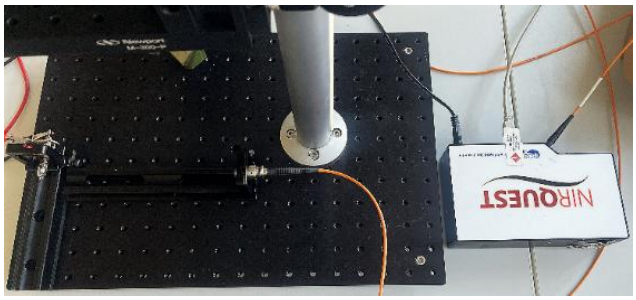


Figure 1 The view on the laboratory set-up used for experimentation.

3.2 Statistical Approach

Using the system response function *H(λ),* the characteristic shape of any measured spectrum can be reconstructed as follows:

$$I_{reconstructed}(T,\lambda) = L_{\lambda}(T_{BBCD}) \cdot H(\lambda) \quad (4)$$

The temperature corresponding to the best-fitting reconstructed spectrum is determined by minimizing the quadratic error between the difference of $I_{reconstructed}(T,\lambda)$and $I_{measured}(\lambda)$. For the purposes of calibration and temperature estimation, two separate datasets are utilized. The system response function is determined using the calibration dataset, while temperature estimation is performed on the test dataset. The results are improved by estimating the temperatures of the calibration dataset and subtracting the resulting MAE from each temperature estimation.

3.3 AI-Based Temperature Prediction Models

To complement the calibration-based measurements, this part of the study focuses on data-driven temperature prediction using artificial-intelligence (AI) models [11-13]. The spectral-to-temperature relationship in high-temperature radiating systems is inherently non-linear and multivariate, making conventional regression models less effective when data are noisy, irregularly sampled, or influenced by complex emissions.

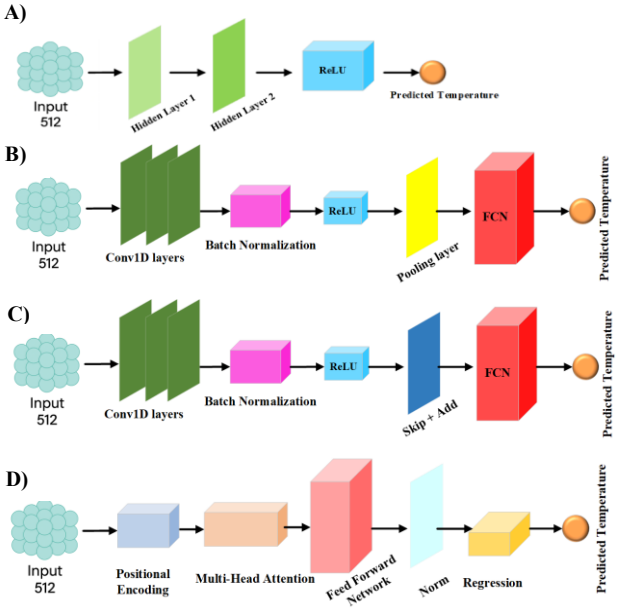


Figure 2 A) FCN architecture diagram. B) 1D-CNN block diagram with convolution and pooling layers. C) Residual block diagram showing skip connection. D) Transformer encoding architecture diagram.

Therefore, we implement and compare four representative deep-learning architectures [14]—Fully Connected Network (FCN), 1D Convolutional Neural Network (CNN1D), Residual Network (ResNet1D), and Transformer Encoder - each designed to capture distinct feature hierarchies within the NIR spectra. Model architectures designed for the study are presented in Figure 2 A-D. These models are trained on both semi-regular and highly irregular datasets (600–1200 °C) to evaluate their ability to generalize under varying sampling conditions. The results provide insight into how architectural choices and learning mechanisms affect robustness, interpretability, and predictive accuracy in portable NIR thermometry.

4 Results

The developed diagnostic pipeline was evaluated on the measured spectral dataset spanning 600–1200°C, where intensity–wavelength curves were collected from irregular temperature intervals. Therefore 80 % of the data were used for training and 20 % for testing to ensure unbiased generalization. The performance of four deep learning models—Fully Connected Network (FCN), 1D Convolutional Neural Network (CNN1D), Residual Network (ResNet1D), and Transformer Encoder—was compared against conventional statistical regression baselines, which yielded notably higher prediction errors. While traditional statistical models struggled to represent the complex, nonlinear relationships between radiative spectral features and temperature, the proposed deep architectures effectively captured spectral dependencies and local emission patterns, resulting in lower Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) – Table 1. This demonstrates the advantage of data-driven learning combined with physics-informed preprocessing, such as baseline correction, smoothing, and normalization.

The dataset comprises NIR emission spectra between 1000 nm and 2200 nm, measured at discrete temperature points from 600-1200°C. The produced irregular and sparsely distributed samples, reflecting realistic industrial monitoring conditions where stable temperature increments cannot always be maintained. To enhance data quality: Asymmetric Least Squares (ALS) baseline correction removed broad background trends. Savitzky–Golay smoothing reduced high-frequency noise. Z-score or percentile normalization ensured intensity comparability across runs.

Among the AI models, the Transformer Encoder achieved the best overall accuracy, followed by FCN, while CNN1D and ResNet1D performed moderately. The FCN showed consistent performance (MAE ≈ 2.4 K), demonstrating its strength in learning global relationships between full-spectrum inputs and temperature. However, its dense architecture lacks spatial awareness of spectral neighbourhoods. The CNN1D and ResNet1D networks leverage convolutional kernels to extract local spectral features, but their performance degraded slightly due to irregular sampling and non-uniform temperature spacing, which weaken the locality assumption of convolutional filters. The Transformer Encoder exhibited the highest robustness to data irregularities. Its self-attention mechanism captures global dependencies between wavelengths, allowing it to dynamically focus on temperature-sensitive spectral regions even when sampling density varies. This explains its superior accuracy and stability across normalization methods.

Two normalization strategies were evaluated: standard-score (z-score) normalization and percentile scaling. As summarized in Table 1, both normalization methods produced comparable overall accuracy trends. However, standard-score normalization yielded slightly better performance for most models, especially for the Transformer Encoder, achieving the lowest MAE (1.97 K) and RMSE

(2.53 K). This suggests that z-score scaling preserved relative spectral dynamics critical for temperature inference, whereas percentile scaling, while more robust to outliers, may have slightly compressed informative radiative contrasts.

The Transformer Encoder, with its attention-based mechanism and positional encoding, remained robust to these irregular patterns, maintaining accuracy even in sparse temperature ranges. Measurement reliability was further influenced by spectral noise, detector saturation, and scattering from oxidation or misalignment. These effects caused baseline drift and distorted emission intensity. To counteract them, an Asymmetric Least Squares (ALS) baseline correction was applied, followed by Savitzky–Golay smoothing, effectively suppressing low-frequency background variations and high-frequency noise while preserving relevant emissivity-driven features.

The results reveal that while statistical models fail to handle complex emission behaviours, deep learning models, particularly the Transformer Encoder, successfully generalize across heterogeneous datasets. The improvement can be attributed to its ability to model non-linear, non-local, and context-dependent spectral interactions that govern thermal emission. The findings emphasize that appropriate normalization, feature-aware model architectures, and physics-informed preprocessing (e.g., BBCD correction) are all critical for reliable temperature prediction under real measurement conditions—especially in portable NIR devices operating under high-radiance and device-dependent transfer functions.

	Normalization method			
	Standard Score		Percentile	
Results(K) / Method	MAE	RMSE	MAE	RMSE
Statistic	-3.15	5.68	-2.83	4.26
FCN	2.43	3.24	2.50	3.24
CNN1D	4.74	6.16	4.28	5.71
ResNet1D	5.09	6.42	4.43	5.53
Transformer Encoder	**1.97**	**2.53**	**2.01**	**2.54**

Table 1 Accuracy comparison of designed models.

5 Conclusions

This study demonstrated the potential of portable NIR spectroscopy for accurate temperature characterization of high-temperature radiating objects in the 600–1200 K range. Through systematic comparison, AI-based models particularly the Transformer and ResNet architectures outperformed classical statistical approaches in accuracy, robustness, and generalization, especially under irregular sampling and noisy measurement conditions. Incorporating blackbody calibration devices (BBCDs) proved essential for maintaining measurement traceability and reducing bias in spectral–temperature mapping. The results highlight that reliable preprocessing, coupled with advanced learning architectures, enables portable NIR systems to achieve laboratory-grade performance in challenging industrial environments.

Acknowledgment

The project was funded by the German Federal Ministry of Education and Research (BMBF) under the grant 16ME0869, titled “Optoelektronische Sensoren für anwendungsnahe Systeme für Lebenswissenschafen und intelligente Fertigung – OASYS-“.

6 References

[1] G. Gullifa, L. Barone, E. Papa, A. Giuffrida, S. Materazzi, and R. Risoluti, ‘Portable NIR spectroscopy: the route to green analytical chemistry’, *Frontiers in Chemistry*, vol. 11, 09 2023.

[2] M. Nobari, N. B. Bouzida, H. Grüger and I. Jablonski, "Indirect measurements with a portable computational NIR spectroscopy - the sources of uncertainty and challenges in a multi-use scenario," *2025 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*, Chemnitz, Germany, 2025.

[3] Xue Q, Yang Y, Ma W, Zhang H, Zhang D, Lan X, Gao L, Zhang J, Tang J. Advances in Miniaturized Computational Spectrometers. Adv Sci (Weinh). 2024 Dec;11(47):e2404448

[4] M. Chirom et al., ‘Non-contact thermometer for measuring surface temperature of photothermal catalysts using near-infrared black-body radiation spectrum’, Device, vol. 2, p. 100467, 07 2024.

[5] T. Pohl *et al.*, ‘Absolute calibration of the spectral responsivity of thermal detectors in the near-infrared (NIR) and mid-infrared (MIR) regions by using blackbody radiation’, *Journal of Sensors and Sensor Systems*.

[6] W. Yang *et al.*, ‘Temperature-Automated Calibration Methods for a Large-Area Blackbody Radiation Source’, *Sensors*, vol. 24, no. 5, 2024.

[7] R. Kang, D. Kyritsis, and P. Liatsis, ‘Comparative analysis of data-driven models for spatially resolved thermometry using emission spectroscopy’, *PLOS ONE*, vol. 20, 01 2025.

[8] Z. Chen, N. Tricard, and S. Deng, ‘Hybrid physics-machine learning model for multispecies and temperature inference from FTIR spectra: Application to ammonia flames’, *Proceedings of the Combustion Institute*, vol. 41, p. 105811, 2025.

[9] J. Xia, Y. Huang, Q. Li, Y. Xiong, and S. Min, ‘Convolutional neural network with near-infrared spectroscopy for plastic discrimination’, *Environmental Chemistry Letters*, vol. 19, 06 2021.

[10] Cui, Ying; Qiu, Kongxin; Gao, Shan; Liu, Hailong; Gao, Rongyan; Chen, Liwei; et al. (2025). Multispectral radiation temperature inversion based on Transformer-LSTM-SVM. Optica Open. Preprint.

[11] X. Mao and Q. Xie, ‘Neural-network-based multispectral thermometry and emissivity reconstruction in cavity high-temperature environments’, *Appl. Opt.*, vol. 64, no. 25, pp. 7304–7314, Sept. 2025.

[12] J. Schneider and M. Vlachos, ‘A Survey of Deep Learning: From Activations to Transformers’, *arXiv [cs.LG]*. 2024.

[13] X. Zhao, L. Wang, Y. Zhang, X. Han, M. Deveci, and M. Parmar, ‘A review of convolutional neural networks in computer vision’, *Artificial Intelligence Review*, vol. 57, no. 4, p. 99, Mar. 2024.

[14] I. D. Mienye and T. G. Swart, ‘A Comprehensive Review of Deep Learning: Architectures, Recent Advances, and Applications’, *Information*, 2024.

DOI 10.5162/iCCC2026/4.6

Halbleiter für Quantumcomptuting: Wertschöpfung und chemischer Fingerabdruck von isotopenreinem Silicium-28

Owen C. Ernst, David Uebel, Roman Brendler, Anne Handschick-Ernst, Jörg Acker
Brandenburgische Technische Universität Cottbus-Senftenberg, Universitätsplatz 1, 01968 Senftenberg

Kurzfassung

Isotopenreine Halbleiter wie Silicium-28 (^{28}Si) nehmen eine Schlüsselposition als Materialplattform für Zukunftstechnologien wie dem Quantencomputing ein. Ihre Wertschöpfungskette ist jedoch komplex und teuer. Da der Anreicherungsprozess in Silicium-28-Tetraflourid ($^{28}SiF_4$) stattfinden muss und dieses industriell nur schwer nutzbar ist, sind chemische Umwandlungsverfahren zu $^{28}SiH_4$ oder andere Silane notwendig. Zwar existiert eine Handvoll chemischer Methoden für die Umwandlung, jedoch besitzen diese individuellen chemische Fingerabdrücke, welche in den weiteren Verarbeitungsschritten zum Quantenbauteil problematisch sind. Die hohen Anforderungen der Halbleiterindustrie und der Quantentechnologie stellt zudem die Analytik vor große Herausforderungen. So kommen gasanalytische Methoden bei der sehr hohen benötigten Reinheit an ihre Nachweisgrenze. Die komplexe Wertschöpfung von ^{28}Si und $^{28}SiH_4$ führen zu extrem hohen Marktpreisen von aktuell über 1‘000 EUR/g. Im ersten Teil dieser Arbeit wird die Wertschöpfungskette für ^{28}Si vorgestellt. Im zweiten Teil werden die Ergebnisse eigener chemischer Umwandlungsexperimente von $^{28}SiF_4$ zu $^{28}SiH_4$ gezeigt. Bereits vor der chemischen Aufreinigung konnte eine sehr gute Gasreinheit von <50 ppm erreicht werden, womit die Ergebnisse aus der Literatur reproduziert und z.T. übertroffen werden könnten.

Abstract

Isotopically pure semiconductors such as silicon-28 (^{28}Si) play a key role as a material platform for future technologies such as quantum computing. However, their value chain is complex and expensive. Since the enrichment process must take place in silicon-28 tetrafluoride ($^{28}SiF_4$), which is difficult to use industrially, chemical conversion processes to $^{28}SiH_4$ or other silanes are necessary. Although there are a handful of chemical methods for conversion, these have individual chemical fingerprints that are problematic in the further processing steps to produce quantum components. The high demands of the semiconductor industry and quantum technology also pose major challenges for analytics. Gas analysis methods, for example, reach their detection limits due to the very high purity required. The complex value chain of ^{28}Si and $^{28}SiH_4$ leads to extremely high market prices of currently over EUR 1,000/g. The first part of this study presents the value chain for ^{28}Si. The second part shows the results of our own chemical conversion experiments from $^{28}SiF_4$ to $^{28}SiH_4$. Even before chemical purification, a very good gas purity of <50 ppm was achieved, thus reproducing and exceeding the results from the literature.

1 Einleitung

Isotopenreine Materialien besitzen veränderte chemische und physikalische Eigenschaften, welche sie für verschiedene technische Anwendungen interessant machen. Als zunehmend relevantes Beispiel ist das ^{28}Si -Isotop zu nennen. ^{28}Si besitzt die höchste jemals gemessene thermische Wärmeleitfähigkeit eines Dielektrikums [1]. Mit 450 W/cmK bei 21 K übersteigt der Wert die Leitfähigkeit von natürlichem Silizium (^{nat}Si) um den Faktor 10. Dies macht den Halbleiter für die Tieftemperaturphysik und mikroelektronische Bauteile mit hoher Wärmeentwicklung attraktiv. Weitere Anwendungen ergeben sich aus der um 58 meV verschobenen Bandlücke [2]. Hochenergetische Messungen zeigen eine präzisere Feinstruktur der nicht-phononischen Photolumineszenzlinien gebundener Exzitonen, die sich bspw. aus B-Dotierung ergeben. Die verbesserte Feinstruktur im Photolumineszenzspektrum kann für Exzitonprozesse in isotopisch reinen Materialien mit hoher Auflösung genutzt werden, bspw. in der Quantensensorik.
Darüber hinaus besitzt ^{28}Si entscheidende Bedeutung als Materialplattform für das Spin-Quantencomputing und die Spintronik [3-5]. Hierfür werden ^{28}Si-Heterostrukturen verwendet, an deren Grenzfläche 2-dimensionale Elektronengase (2DEG) oder Lochgase (2DHG) entstehen, in denen Landungsträgerspinzuständen vereinzelt, manipuliert, gemessen und schlussendlich als Qubits verwendet werden können. Die Isotopenreinheit der Materialplattform ist entscheidend, da Kernspins von ungeradzahligen Isotopen wie ^{29}Si mit den Landungsträgerspins über Fermi-Kontakt-Wechselwirkung interagieren und die Dekohärenzzeit maßgeblich senken. Geradzahlige Isotope wie ^{28}Si sind dagegen Kernspin-frei (I=0). Man spricht auch vom „Spin-Vakuum“. Bei einer Absenkung des ^{29}Si-Gehalts von 100 % auf 1 ppm kann eine Dekohärenzzeitsteigerung von T_2=20 µs auf T_2=60 s erfolgen [6].
Silicium-Spin-Qubits auf 300-mm-Wafern zeigen bereits heute hohe Fidelity-Werte von >99% (two-qubit control fidelity and measurement fidelity), wodurch die prinzipielle Skalierbarkeit des Ansatzes gezeigt ist [7]. Die Kohärenz der Systeme ist jedoch auch heute auf T_2^{echo}≈100-800 ms begrenzt [6-9]. Grund hierfür ist die geringe aktuelle Materialqualität von isotopenreinem ^{28}Si.
Die Wertschöpfungskette für ^{28}Si ist komplex und teuer. Sie erfordert aufwendige Anreicherungs- und chemische Umwandlungsverfahren, die einen chemischen Fingerabdruck hinterlassen. Die einhergehenden Verunreinigungen

DOI 10.5162/iCCC2026/4.6

senken die Kohärenzzeit der Ladungsträger [6,10]: T_2^{echo}-Zeit sinkt bei Donor-Konzentrationen $C_E=1.2\times10^{14}$ /cm^3 bei 1 ppm ^{29}Si von 60 s auf 200 ms ab. Es fehlen instrumentelle Methoden zur kostengünstigen Aufreinigung und Sicherstellung der chemischen und isotopischen Reinheit.
Bis 2022 wurde Material über individuelle F&E-Kanäle aus Russland beschafft. Aufgrund der geopolitischen Spannungen und Handelsrestriktionen stellt sich seitdem ein Liefer- und Materialengpass ein. Für industrielle Hersteller sind die schlechte Verfügbarkeit und die eingeschränkte Materialqualität ein technisches und wirtschaftliches Problem. Nichtsdestotrotz wird das wenige verfügbare Material für hohe Preise > 1‘000 EUR/g gehandelt.
Die limitierte Verfügbarkeit von ^{28}Si wird von der Transatlantic Quantum Community (TQC) seit 2025 als eine „key vulnarability“ eingestuft. Der Bericht *Critical Vulnerabilities in the Quantum Computing Supply Chain within the NATO Alliance* (2025) empfiehlt “[to] Develop an under-standing of the key requirements needed to support the fabrication of semiconductor spin systems, including any specialized requirements for the successful integration of isotopically pure ^{28}Si into the process flow”.
In dieser Arbeit geben wir einen Überblick über die komplexe Wertschöpfungskette für isotopenreine Halbleiter und die bisherigen Synthesemethoden. Darüber hinaus wird eine neue Methode zur lokal selektiven, laser-gestützten Prüfabscheidung vorgestellt, mit der die chemische und isotopische Reinheit von ^{28}Si bereits im Präkursorgas ohne großen Materialverlust zugänglich gemacht wird.
Dies eröffnet erstmals die Möglichkeit zu Produktion von reproduzierbar electronic grade ^{28}Si, welche die Anforderungen der Halbleiterindustrie erfüllt.

2 Technischer Überblick

2.1 Wertschöpfungskette

2.1.1 Isotopische Anreicherung

Es existiert eine Bandbreite von physikalischen Methoden zur isotopischen Anreicherung [11]. Wirtschaftlich konnten sich bisher nur Gegenstromzentrifugen durchsetzen.
In Gaszentrifugen wird die Isotopenanreicherung durch die Nutzung der massenabhängigen Trägheit von Molekülen erreicht. Die Isotopentrennung findet in rotierenden Zylindern statt. Ein in die Zylinder eingeleitetes Gas, z. B. SiF_4, wird einer permanenten Beschleunigung ausgesetzt. Die resultierende Trägheitskraft F kann anhand der folgenden einfachen Beziehung

$$F=Mv^2r^{-1}$$

(M – Molmasse; v – Umlaufgeschwindigkeit; r – Radius des Zylinders)
berechnet werden und sorgt für die Isotopentrennung.
Die Anreicherung wird in der Regel in SiF_4 durchgeführt, da es als eine von wenigen Si-Verbindungen die chemischen und physikalischen Anforderungen erfüllt [12,13]:

i. Der Dampfdruck der Substanz muss 5−10 mmHg (≙ 0,7–1,3 kPa) überschreiten, da die Anreicherung eine stabile gasförmige Quelle erfordert.
ii. Für eine ausreichende Anreicherung darf die Molekülmasse der Moleküle nicht weniger als 40 u betragen. Höhere Molekülmassen sind in der Regel mit höheren Anreicherungsraten und einem geringeren Energieverbrauch verbunden.
iii. Die anzureichernde Substanz muss eine ausreichend hohe thermische Stabilität aufweisen.
iv. Es sollte keine chemische Reaktion zwischen den Molekülen anzureichernden Substanz oder mit Anreicherungskammer stattfinden.

Moderne Gaszentrifugen, auch bekannt als ZIPPE- oder KAMENEV-Zentrifugen, bestehen aus hohen, vertikal rotierenden Zylindern. In der Regel bestehen sie aus kohlenstofffaserverstärktem Stahl mit einem Zylinderradius von 5 cm und können Drehzahlen von 750 ms−1 und mehr erreichen [14]. Hohe Drehzahlen werden nicht mehr durch einfache Elektromotoren erreicht, sondern durch ein rotierendes Magnetfeld, das den Zylinder selbst in Drehung versetzt. Auf ein Molekül SiF_4 wirkt eine Kraft von bis zu 12.000 eVm (>1.000.000 g). Dem wirkt ein Druckgefälle zwischen dem gegen die Zylinderwand komprimierten Gas und dem an der Drehachse erzeugten Vakuum entgegen. Ist der Druck an der Wand aufgrund hoher Umwälzgeschwindigkeiten zu hoch, kann das Gas an den Wänden der Zentrifuge kondensieren.
In der Gegenstromzentrifuge entsteht zusätzlich zum Gasstrom von innen nach außen ein vertikaler Druckunterschied, der die Moleküle nahe der Drehachse nach unten (oder oben) drückt. Durch Erhitzen des Zylinderkopfes und Veränderung der Geometrie des Zylinderinneren kann ein Gegenstrom erzeugt werden, der die inneren Druckunterschiede aufrechterhält oder sogar verstärkt. Dieser Strom führt zu horizontalen Zonen innerhalb des Zylinders. In jeder Zone findet eine horizontale Anreicherung von innen (leichte Moleküle) nach außen (schwere Moleküle) statt. Aufgrund des vertikalen Druckunterschieds werden die schweren Moleküle langsam in die darüber (oder darunter) liegende Zone abgegeben, während die leichten Moleküle in die entgegengesetzt liegende Zone wandern. Der Anreicherungsgrad steigt mit jeder zusätzlichen Zone. Daraus ergeben sich zwei Strategien. Zum einen können viele kurze Zentrifugen miteinander verbunden werden. Der Anreicherungsgrad kann durch die Reihenschaltung der Zentrifugen erhöht werden, und der Durchsatz kann durch Parallelschaltung gesteigert werden. Diese technische Lösung ist sehr flexibel und lässt sich schnell an veränderte Systemparameter oder Produktanforderungen anpassen. Die zweite Strategie basiert auf einer geringen Anzahl hoher Zentrifugen. Dadurch lassen sich viele Anreicherungszonen innerhalb eines Zylinders realisieren. Dadurch läuft der Prozess schneller und erfordert weniger wartungsintensive Einzelkomponenten. Außerdem wird die Kontamination durch Zylinderwände und Gasleitungen minimiert. Die Herausforderung bei dieser Methode ist die Stabilität der hohen Zylinder. Beim Anlaufen der Rotationsmotoren müssen die Biegefrequenzen der Zylinder überwunden werden. Zylinder, die länger als 1 m sind, laufen Gefahr, sich kritisch zu verformen, wenn sie nicht durch spezielle Federkonstruktionen stabilisiert werden.

DOI 10.5162/iCCC2026/4.6

Urenco Ltd. ist der weltweit größte Hersteller von isotopisch reinen Materialien und setzt dabei ausschließlich auf Gegenstromgaszentrifugen. Nur die Anreicherungsanlagen der russischen TVEL Fuel Company, einer Tochtergesellschaft der staatlichen Atomenergoprom-Gruppe, verfügen über eine ähnliche Produktionskapazität.
Sowohl Urenco als auch TVEL haben nachgewiesen, dass sie $^{28}SiF_4$ auf eine isotopische Reinheit von bis zu 99,9999 % anreichern können.

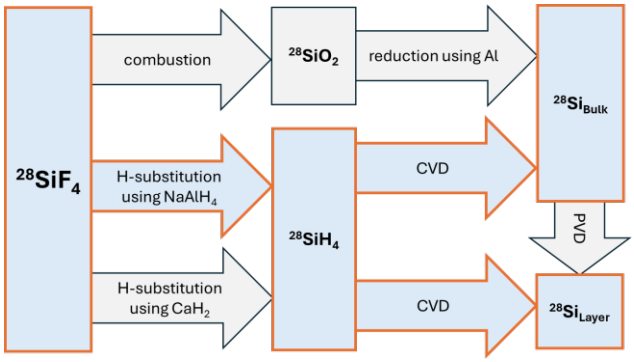


Bild 1 Umwandlungsrouen von $^{28}SiF_4$.

2.1.2 Chemische Umwandlung

$^{28}SiF_4$ ist für die meisten industriellen Prozesse ungeeignet [11]. Der Grund liegt u.a. in der außergewöhnlich hohen negativen Standardbildungsenthalpie (<-1500 kJ/mol [15]) des Moleküls. Die Si-F-Bindung gilt als die stärkste bisher gemessene Einfachbindung. Industriell benötigt werden verschiedene Silan-Spezies wie Monosilan SiH_4 oder Trichlorsilan $HSiCl_3$ als CVD-Präkursoren.
Die Umwandlung von inerten Fluoriden in isotopisch reine Halbleiterpräkursoren stellt eine besondere Herausforderung dar (**Bild 1**). Die chemische Reduktion von Siliciumtetrafluorid wird seit den 1960er Jahren als Verfahren zur Herstellung von SiH4 untersucht.
Möglich ist die quantitative Umwandlung von Halbmetallhalogeniden (MX_y) in ihre Wasserstoffverbindungen (MH_y) unter Verwendung von Hydriden (AH_x).

$$MX_y + AH_x \rightarrow MH_y + AX_x$$

In frühen Forschungsarbeiten wurden Umwandlungen mit gelösten Alkalihydriden durchgeführt, die Aluminium (Al), Lithium (Li), Natrium (Na) oder Kombinationen davon enthielten. Diese Hydride müssen vorhergehend in organischen Lösungsmitteln wie Dimethylether, Diethylenglykol oder Tetrahydrofuran gelöst werden, wodurch C-Verunreinigungen enstehen [16,17].
Alternativ kann festes Calciumhydrid (CaH_2) zur Reduktion verwendet wurde. Der Vorteil dieser Methode liegt in ihrer Einfachheit, da sie nur $^{28}SiF_4$ und CaH_2 umfasst und somit die Einführung kritischer Verunreinigungen, insbesondere C und Al, vermeidet.

$$SiF_4 + 2\ CaH_2 \rightarrow SiH_4 + CaF_2$$

2.1.3 Abscheidung

Für die Synthese von ^{28}Si-Volumenfestkörpern wird die $^{28}SiH_4$ in angepassten Siemens-Ofen thermisch zersetzt. Die Reaktionsbedingungen wie Temperatur und Druck sind entscheidend für die Qualität und die Struktur der gewachsenen Schicht und ermöglichen die Bildung von polykristallinem Silizium [18]. ^{28}Si-Einkristalle können im Anschluss bspw. über Zonenschmelzverfahren erhalten werden. Diese Methode wurde zur Herstellung größerer Mengen ^{28}Si für das Avogadro-Projekt genutzt [19,20]. Bemerkenswert ist, dass zuvor mit dem CaH_2-Verfahren erhebliche Mengen (5 kg) an $^{28}SiH_4$ hergestellt wurden.
Für Quantensysteme wird ^{28}Si als dünne Schichten in Heterostack benötigt. $^{28}SiH_4$ wird hierbei als Präkursor für Niederdruck-CVD-Prozesse verwendet. Epitaktisches Wachstum ist bei 650 °C und einem Druck von 20 mbar möglich, wodurch sich eine Schicht aus ^{28}Si auf einem Siliziumwafer mit natürlicher Isotopenzusammensetzung bilden kann. Mit diesem Verfahren können Epitaxieschichten mit einem Durchmesser von über 300 mm hergestellt werden [7,21].

2.2 Chemische Reinheit

Die Anreicherung von $^{28}SiF_4$ ist bereits heute industriell möglich und Material auf dem Weltmarkt für Preise <500 EUR/g erhältlich. Herausforderung ist die chemische Umwandlung des Fluorids zu $^{28}SiH_4$ und ^{28}Si [11].
Erste Versuche begannen 1958 in den USA [22]. Die Hauptstrategie bestand in der Umwandlung von $^{28}SiF_4$ zu $^{28}SiO_2$ und der anschließenden Reduktion mit Al-Pulver bei 1100 °C

$$^{28}SiF_4 + O_2 \rightarrow {}^{28}SiO_2 + 2\ F_2$$
$$3\ {}^{28}SiO_2 + 4\ Al \rightarrow 3\ {}^{28}Si + 2\ Al_2O_3$$

Die Methode wurde um 1994 in einer amerikanisch-russisch-deutschen Kooperation wiederbelebt.[20] $^{28}SiO_2$ stammte aus dem Oak Ridge National Laboratory (ORNL) und dem russischen Nationalen Metrologieinstitut (NMI). Bemerkenswert ist, dass das NMI-Material aufgrund hoher Verunreinigungen unbekannter Herkunft verworfen wurde, während das ORNL-Material in Zusammenarbeit zwischen der Physikalisch-Technischen Bundesanstalt (PTB) und Wacker-Chemitronic in Deutschland mit der oben beschriebenen Al-Methode reduziert wurde. Die resultierenden ^{28}Si-Kristalle mit einer Masse von 300 g wiesen hohe Verunreinigungswerte mit C ($1{,}3 \times 10^{17}$ cm^{-3}), O ($1{,}1 \times 10^{16}$ cm^{-3}) und B (10^{18} cm^{-3}) auf.
Bei Experimenten Ende der 1990er Jahre, wurde die Reduktion mit Lithiumaluminiumhydrid ($LiAlH_4$) oder Natriumaluminiumhydrid ($NaAlH_4$) genutzt. Die dabei entstehenden Fluorsalze verschlammen jedoch das Lösungsmittel, was die Umsetzung und Skalierung dieser Verfahren technisch erschwert. Darüber hinaus führt das Lösungsmittel organische Lösungsmittel ein, die als Kohlenstoffquellen – elektrisch aktive Donatoren in Silizium – wirken können. Ebenfalls wirkt Al im Si elektrisch aktiv, was eine aufwendige Reinigung erforderlich macht. Es wurden Ausbeuten von über 99 % berichtet [16].
Fortschritte wurden mit einem damals neuartigen Ansatz erzielt, bei dem festes Calciumhydrid (CaH_2) zur Reduktion verwendet wurde. Da das verwendete $^{28}SiF_4$ aufwendig in der Beschaffung und kostspielig ist, spielt die Effizienz eine wichtige Rolle – d. h. die Umwandlung von ^{28}Si aus dem Rohmaterial in das Zielmaterial. Für den CaH_2-

DOI 10.5162/iCCC2026/4.6

Ansatz wurden Wirkungsgrade von bis zu 90 % angegeben. Voltaix, ein amerikanisches Chemieunternehmen, das das CaH_2-Verfahren einsetzt, erreicht jedoch nur Wirkungsgrade von 66 %.
Reinheitswerte werden in den meisten Fällen nur für abgeschiedenen ^{28}Si-Festkörper (dünne Schichten oder Volumenkristalle) veröffentlicht. Die im Festkörper anwendbaren Charakterisierungsmethoden wie SIMS und FTIR besitzen i.d.R. um Größenordnungen bessere Nachweisgrenzen als Gasanlysemethode. Jedoch kann die im Festkörper gemessene chemische Reinheit nur bedingt als Maß für den chemischen Fingerabdruck der chemischen Umwandlung zu $^{28}SiH_4$ dienen, da die Abscheidung des Gases einen hohen Aufreinigungseffekt besitzt.
Die wenigen Studie, die Gas- und Festkörperreinheit vorstellen, zeigen, dass Gasverunreinigungen von 100 ppm nur noch mit <1 ppm im Festkörper auftreten.

3 Ergebnisse & Diskussion

In eigenen Versuchen wurden die Ergebnisse aus der Literatur reproduziert. Durch Reduktion mit Alkalihydriden konnte $^{28}SiH_4$ mit Ausbeute von >90% dargestellt werden. Ausbeute und Selektivität der Reaktion ließen sich mithilfe von zwei Maßnahmen verbessern:

i. Die Reaktion muss stark gekühlt (<-20°C) werden, um der exothermen Natur entgegenzuwirken,
ii. Die Reaktionslösung muss stark verdünnt werden (<0.2 M), um Verschlammung zu verhindern.

Diese Maßnahmen führen jedoch auch zu einer signifikanten Erhöhung der Reaktionszeit (> 7 d) und einer schlechteren Skalierbarkeit des Prozesses.

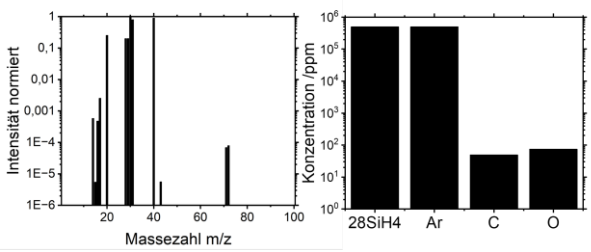


Bild 2 Gas-MS-Spektren der $^{28}SiH_4$-Proben.

Bereits ohne die Nutzung der etablierten Aufreinigungsverfahren der Reinstgasindustrie, konnten mit dieser Methode gute Verunreinigungswerte von C (50 ppm ≙ 5×10^{17} cm^{-3}) und O (100 ppm ≙ 1×10^{18} cm^{-3}) erhalten werden (**Bild 2**). Aufgrund der zusätzlich aufreinigenden Wirkung der Abscheidungsverfahren, sind resultierende Festkörperwerte im ≤ ppm-Bereich zu erwarten.
Eine Herausforderung stellt die Quantifizierung der Gasqualität nach Reinigung dar. Herkömmliche Messmethoden wie Gas-Massenspektrometrie stoßen bei Konzentrationen < 1 ppm an ihre Nachweisgrenzen. Für CVD-Prozesse spielt nicht nur die Konzentration an Verunreigungen eine Rolle, sondern auch wie effektiv sich diese während des CVD-Prozesses abscheiden. So bringt CO_2 als C-Quelle weniger kritische Verunreinigungen ein als CH_4, da die Abscheiderate des CH_4 um ein Vielfaches höher ist.
Um niedrigere Nachweisgrenzen im ppb-Bereich oder besser zu erreichen, werden Prüfabscheidungen durchgeführt, bei denen ein Teil der Präkursoren nach der Synthese in spezialisierten CVD-Anlagen zu Feststoff abgeschieden wird. Dieser Feststoff soll mittels Feststoffanalytik, etwa Sekundär-Ionen-Massenspektrometrie (SIMS), auf Reinheit geprüft werden.

4 Fazit & Ausblick

Im labortechnischen Aufbau wurde hochreines $^{28}SiH_4$-Gas erzeugt, welches abgeschiedene Festkörper mit niedrigen Verunreinigungen im ppm Bereich verspricht. Diese Reinheit ist mit dem Literaturvergleich im Einklang, allerdings noch nicht genügend hoch für die direkte Verarbeitung zu Quantencomputern. Daher müssen Verbesserungen sowohl in der Prozesstechnik als auch in der Analytik vorangetrieben werden.
In der Halbleiter- und Reinstgastechnik haben sich verschiedene Aufreinigungsverfahren etabliert. Diese müssen prozesstechnisch an den chemischen Fingerabdruck der ^{28}Si-Wertschöpfungskette angepasst werden, um chemische Reinheiten von << 1 ppm und isotopische Reinheiten von > 5N zu gewährleisten.
Die resultierende Reinheit liegt unter der Nachweisgrenze der meisten gasanalytischen Methoden wie Gas-Massenspektrometer. Es werden innovative, materialsparende Methoden benötigt mit geringeren Nachweisgrenzen benötigt. Dies betrifft insbesondere CVD-Prüfabscheidung zum Festkörper und folgendes HR-SIMS. Selektive Laserabscheidung zu schnellen und materialsparenden Prüfabscheidung befindet sich an der BTU Cottbus-Senftenberg bereits in der Entwicklung.

5 Literatur

[1] A. V Inyushkin et al., J Appl Phys, vol. 123, no. 9, p. 095112, 2018
[2] D. Karaiskaj et al., Phys Rev Lett, vol. 86, no. 26, pp. 6010–6013, 2001
[3] B. E. Kane, Nature, vol. 393, no. 6681, pp. 133–137, 1998
[4] O. C. Ernst et al., Zeitschrift für Kristallographie-Crystalline Materials, vol. 237, no. 4–5, pp. 191–200, 2022.
[5] Y. Liu et al., J Cryst Growth, vol. 593, p. 126733, 2022
[6] W. M. Witzel et al., "Phys Rev Lett, vol. 105, no. 18, p. 187602, 2010.
[7] P. Steinacker et al., Nature, vol. 646, no. 8083, pp. 81–87, Oct. 2025,
[8] G. Scappucci et al., MRS Bull, vol. 46, no. 7, pp. 596–606, 2021
[9] A. M. J. Zwerver et al., "Nat Electron, vol. 5, no. 3, pp. 184–190, 2022
[10] K. M. Itoh and H. Watanabe, MRS Commun, vol. 4, no. 4, pp. 143–157, 2014
[11] O. C. Ernst et al., Molecules, vol. 29, no. 17, p. 4222, 2024.
[12] E. V. Levin et al., Atomic Energy, vol. 77, no. 4, pp. 760–767, 1994
[13] V. D. Borisevich and E. V. Levin, Sep Sci Technol, vol. 36, no. 8–9, pp. 1697–1735, 2001,
[14] D. Kushner, IEEE Spectr, vol. 50, no. 3, pp. 48–53, 2013,
[15] P. Sukkaew, et al., ECS Journal of Solid State Science and Technology, vol. 5, no. 2, pp. P27–P35, 2016
[16] P. G. Sennikov et al., J Radioanal Nucl Chem, vol. 306, no. 1, pp. 21–30, 2015
[17] G. G. Devyatykh et al., Doklady Chemistry, vol. 391, no. 4/6, pp. 204–205, 2003
[18] W. L. M. Weerts, et al., J Electrochem Soc, vol. 145, no. 4, p. 1318, 1998.
[19] E. W. Becker et al., Atomwirtschaft, vol. 11, 1979.
[20] P. Becker et al., Meas Sci Technol, vol. 20, no. 9, p. 092002, 2009,
[21] V. Mazzocchi et al., J Cryst Growth, vol. 509, pp. 1–7, 2019
[22] J. P. Gordon and K. D. Bowers, Phys Rev Lett, vol. 1, no. 10, pp. 368–370, 1958

DOI 10.5162/iCCC2026/4.7

Mehr als Bionik - Biofunktionelle Polymere als innovative neue Materialien

More than bionics - biofunctional polymers as innovative new materials

Sophia Rosencrantz, Fraunhofer IAP, Potsdam, Deutschland
Joachim Storsberg, Fraunhofer IAP, Potsdam, Deutschland
Ruben R. Rosencrantz, BTU Cottbus-Senftenberg, Senftenberg, Deutschland; Fraunhofer IAP, Potsdam, Deutschland

Kurzfassung

Biofunktionelle Polymere integrieren biologische Funktionsprinzipien in synthetische Materialien und schaffen damit eine neue Materialklasse mit gezielt einstellbaren Eigenschaften. Durch die Integration von Zuckermolekülen (Glykanen), Enzymen oder biomimetischen Strukturen entstehen Hochleistungspolymere mit einzigartigen Anwendungsperspektiven in Medizintechnik, Sensorik und funktionalen Beschichtungen. Der Beitrag zeigt aktuelle Forschungsergebnisse zu Glykopolymeren, wirkstofffreisetzenden Kontaktlinsen, 3D-biogedruckten Implantaten und biofunktionalisierten Bulk-Materialien. Diese Materialien gehen über reine Bionik hinaus: Sie kombinieren synthetische Polymerchemie mit molekularbiologischen Funktionsmechanismen und eröffnen disruptive Innovationspotenziale für die Materialindustrie.

Abstract

Biofunctional polymers integrate biological functional principles into synthetic materials, creating a new class of materials with precisely tunable properties. By incorporating sugar molecules (glycans), enzymes or biomimetic structures, high-performance polymers are created with unique application perspectives in medical technology, sensor technology and functional coatings. This contribution presents current research results on glycopolymers, drug-eluting contact lenses, 3D-bioprinted implants and biofunctionalized bulk materials. These materials go beyond pure bionics: they combine synthetic polymer chemistry with molecular biological functional mechanisms and open up disruptive innovation potentials for the materials industry.

1 Von der Natur lernen - mehr als Bionik

Die Bionik kopiert Strukturen und Mechanismen der Natur für technische Anwendungen, vom Lotuseffekt bis zum Klettverschluss. Biofunktionelle Polymere gehen jedoch einen entscheidenden Schritt weiter: Sie integrieren die molekularen Erkennungs- und Wechselwirkungsprinzipien biologischer Systeme direkt in synthetische Materialien. Dabei werden natürliche Funktionsbausteine wie Zucker (Glykane), Peptide oder Enzyme mit Polymeren kombiniert, um Materialien mit völlig neuen Eigenschaftsprofilen zu schaffen. Biofunktionelle Materialien sind definiert als Materialien, die aus Biomasse hergestellt sind und/oder mit Biomasse wechselwirken. Diese Definition umfasst sowohl die stoffliche Herkunft als auch die funktionelle Interaktion auf molekularer Ebene. Die biologische Wechselwirkung kann dabei auf verschiedenen Ebenen stattfinden: von der Zellerkennung über Proteinbindung bis hin zu enzymatischer Katalyse.

1.1 Glykopolymere: Synthetische Nachbildung natürlicher Mucine

Glykane sind die komplexeste und vielfältigste Klasse von Biomolekülen. Sie bilden die Glykokalyx, eine dichte Zuckerschicht auf allen menschlichen Zellen, und sind verantwortlich für fundamentale biologische Prozesse wie Zell-Zell-Kommunikation, Immunantworten und Pathogenabwehr. Zuckerstrukturen sind im menschlichen Körper allgegenwärtig: Sie umgeben die Zellen wie ein Gerüst, das sie einerseits schützt und andererseits Interaktionen mit anderen Kompartimenten ermöglicht. Die sogenannten Glykane treten als Glykolipide, Proteoglykane oder Glykoproteine auf.

Mucine, die natürlichen „Schleimproteine“, bestehen aus einem Proteinrückgrat mit dicht angebundenen Zuckerketten und sind für ihre außergewöhnlichen tribologischen Eigenschaften sowie ihre Schutz- und Lubrikationsfunktion bekannt. Die strukturelle Komplexität der Oligosaccharide ermöglicht es ihnen, als Informationsträger zu fungieren – viele Proteine besitzen spezifische Zuckerstrukturen, die entscheidend zur Funktion des Proteins beitragen.

Glykopolymere sind synthetische Polymere mit kovalent gebundenen Zuckerstrukturen. Sie imitieren die multivalente Präsentation von Glykanen auf Mucinen, ersetzen aber das Proteinrückgrat durch ein chemisch stabiles, synthetisches Polymerrückgrat. Zwei komplementäre Synthesewege sind etabliert: die Polymerisation von Glykomonomeren und die post-polymerisations-Funktionalisierung bestehender Polymere mit Zuckern. Durch chemische Sulfatierung können anionische, kationische oder zwitterioni-

DOI 10.5162/iCCC2026/4.7

sche Glykopolymere mit gezielt einstellbaren Eigenschaften herstellen. Neuste Arbeiten zu sulfatierten Glykopolymeren zeigen, dass diese Strukturen als hervorragende Mucin-Mimetika dienen können.
Diese Glykopolymere zeigen außergewöhnliche Eigenschaften, die sie für industrielle Anwendungen attraktiv machen: Sie sind vollständig biokompatibel und zeigen ausgezeichnete Verträglichkeit mit Zellen und Geweben. Ihre tribologischen Eigenschaften übertreffen in vergleichenden Tests sogar die der kommerziell eingesetzten Hyaluronsäure. Die multivalente Präsentation der Zuckerstrukturen ermöglicht die spezifische Erkennung von Lektinen und anderen glykan-bindenden Proteinen, ein Mechanismus, der für Diagnostik und gezielte Wirkstofffreisetzung genutzt werden kann. Gleichzeitig bieten synthetische Glykopolymere eine deutlich höhere chemische und thermische Stabilität als natürliche Mucine, was ihre Verarbeitung und Lagerung erheblich vereinfacht.

2 Biofunktionelle Kontaktlinsen

2.1 Wirkstofffreisetzende Kontaktlinsen mit Glykopolymerbeschichtung

Ein herausragendes Anwendungsbeispiel für biofunktionelle Polymere sind drug-eluting Kontaktlinsen. Bei der konventionellen topischen Applikation von Augentropfen werden weniger als 5% des Wirkstoffs tatsächlich vom Auge aufgenommen, der überwiegende Teil wird durch Tränenfluss und Lidschlag weggewaschen oder systemisch absorbiert. Gleichzeitig klagen etwa 30% der Kontaktlinsenträger über Augenirritationen, die häufig durch Reibung zwischen Linse und Augenoberfläche verursacht werden. Hier kann ein biofunktionelles Liposom-Polymer-System Abhilfe schaffen, das beide Probleme simultan adressiert.
Die Kontaktlinse wird mittels Layer-by-Layer (LbL) Selbstorganisation mit Glykopolymeren beschichtet. Bei diesem Verfahren werden abwechselnd entgegengesetzt geladene Glykopolymer-Elektrolyte auf die Oberfläche aufgebracht, wobei jede Schicht durch elektrostatische Wechselwirkungen stabilisiert wird. Die Synthese der Glykopolymer-Elektrolyte erfolgt schutzgruppenfrei und in wässriger Lösung, um möglichst wenige Syntheseschritte durchführen zu müssen.
Diese mucin-mimetischen Beschichtungen reduzieren die Reibung mit dem Auge signifikant, in tribologischen Tests zeigen diese Polymere bessere Ergebnisse als die kommerziell eingesetzte Hyaluronsäure. Der Mechanismus basiert auf der hydrophilen, stark hydratisierten Natur der Zuckerketten, die eine Wasserschicht an der Oberfläche binden und so die Reibung minimieren.
In die Glykopolymer-Matrix können zusätzlich wirkstoffgefüllte Liposomen eingebettet werden. Whispering Gallery Mode (WGM) Spektroskopie und Quartz Crystal Microbalance mit Dissipationsmonitoring (QCM-D) Messungen bestätigen, dass die Liposomen in der Beschichtung intakt bleiben. Die adsorbierte Masse, pH-Stabilität und Integrität der Glykopolymer-Beschichtungen sowie der eingebetteten Liposomen wurden durch diese markierungsfreien Charakterisierungsmethoden verifiziert. Die Glykopolymer-Dünnschichten mit und ohne Liposomen sind im physiologischen pH-Bereich stabil. Die Liposomen ermöglichen eine gleichmäßige Wirkstofffreisetzung über längere Zeiträume.
Dieses System vereint somit drei therapeutische Vorteile: verlängerte Wirkstofffreisetzung für verbesserte Bioverfügbarkeit, erhöhter Tragekomfort durch reduzierte Reibung, und hohe Biokompatibilität durch die naturinspirierte Glykopolymer-Chemie.

3 Enzyme als Additive in der Kunststoffverarbeitung

Proteine und Enzyme als therapeutische Wirkstoffe reagieren empfindlich auf Temperaturschwankungen, pH-Änderungen und mechanische Belastungen. Biofunktionelle Ansätze können zur Stabilisierung von Biomolekülen in verschiedenen Formulierungen eingesetzt werden..

3.1 Immobilisierung und Formulierung für industrielle Anwendungen

Für industrielle Anwendungen wurden systematisch Methoden entwickelt, um Enzyme so zu formulieren, dass sie auch extreme Verarbeitungsbedingungen überstehen. Dies ist besonders relevant für die Integration von Enzymen in Kunststoffe, aber auch für die Entwicklung stabiler industrieller Enzymformulierungen.
Die Adsorption von Enzymen an anorganischen Trägern mit großer Oberfläche hat sich als besonders effektive Stabilisierungsmethode erwiesen. Trägersubstanzen wie Kaolin, Siliziumdioxid oder Zeolithe bieten mehrere Vorteile: Sie sind günstig und in großen Mengen verfügbar, erfordern keine kovalente Modifikation des Enzyms, und ihr Einbau in Polymere ist ein Standardverfahren der Kunststoffverarbeitung. Die Adsorption an diese Träger stabilisiert Enzyme gegen Temperatur, pH-Änderungen und mechanische Scherung. In umfangreichen Studien wurde gezeigt, dass so stabilisierte Enzyme die thermoplastische Kunststoffverarbeitung bei Temperaturen bis zu 160°C überstehen können, und enzymatisch aktiv bleiben. Die thermische Stabilität der Enzyme wird durch die Einbettung in den Kunststoff sogar um etwa 40 K gesteigert, wahrscheinlich, weil der Kunststoff zunächst thermische Energie absorbiert, bevor diese auf das Enzym einwirken kann.
Diese Erkenntnisse sind nicht nur für biofunktionalisierte Kunststoffe relevant, sondern auch für die Formulierung von Enzympräparaten für industrielle Prozesse, Waschmittel, Lebensmittelverarbeitung und andere Anwendungen, bei denen Enzyme unter herausfordernden Bedingungen stabil bleiben müssen. Insbesondere von Interesse ist aber die Tatsache, dass auf diese Weise depolymerisierende Enzyme in Kunststoffe eingearbeitet werden können und so eine Art Selbstzerstörungsmechanismus integriert werden kann. Die biologische Abbaubarkeit solcher enzymadditivierter Polymere kann deutlich verbessert werden.

DOI 10.5162/iCCC2026/4.7

4 3D-Biodruck von Knorpelimplantaten

Knorpelschäden haben ein geringes Heilungspotenzial, da Knorpelgewebe nicht durchblutet ist und somit keine regenerativen Stammzellen oder Wachstumsfaktoren über den Blutkreislauf erreichen können. Unbehandelte Knorpeldefekte führen häufig zu Arthrose, einer degenerativen Gelenkerkrankung, für die es derzeit keine effektive kausale Therapie gibt. Um personalisiert Knorpelimplantate herzustellen, ist der 3D-Biodruck prädesitiniert. Allerdings müssen hierfür neue Materialien entwickelt werden, die der natürlichen Umgebung der Knorpelzellen nahekommt. Auch hier können biofunktionalisierte Polymere helfen.

4.1 Glykopolymer-basierte Biotinten

Neuartige Biotinten enthalten körpereigene Knorpelzellen (Chondrozyten) der Patienten, die aus einer kleinen Gewebeprobe isoliert und im Labor vermehrt werden. Diese Zellen werden in ein Glykopolymer-Hydrogel eingebettet, das als druckbare Biotinte dient. Glykopolymere sind für diese Anwendung besonders geeignet, da sie die natürliche Umgebung der Knorpelzellen, die extrazelluläre Matrix des Knorpels, biomimetisch nachbilden können
Die extrazelluläre Matrix des Knorpels besteht zu einem wesentlichen Teil aus Proteoglykanen, die ein Proteinrückgrat mit dicht angebundenen Glykosaminoglykanen (sulfatierten Zuckerketten) besitzen. Synthetischen Glykopolymere können diese Struktur nachahmen: Die Zuckerketten interagieren mit den Zelloberflächenrezeptoren der Chondrozyten und fördern Zelladhäsion, Proliferation und die Aufrechterhaltung des chondrogenen Phänotyps. Im Gegensatz zu vielen anderen Hydrogel-Systemen, in denen Chondrozyten zu Fibroblasten-ähnlichen Zellen dedifferenzieren, könnten Glykopolymer-Matrices die Produktion von knorpelspezifischen Matrixproteinen wie Kollagen Typ II und Aggrecan unterstützen.
Die Biotinten können während oder nach dem Druck durch verschiedene Mechanismen vernetzt werden: photochemisch, enzymatisch oder ionisch. So kann die gewünschte Form und mechanische Stabilität erzeugt werden. Die Vernetzungsdichte und damit die mechanischen Eigenschaften lassen sich präzise einstellen, um den Anforderungen an die Implantatstelle gerecht zu werden.

4.1 Personalisierte Implantate und klinische Perspektiven

Der 3D-Biodruck ermöglicht es, patientenspezifische Implantate mit exakt angepasster Geometrie herzustellen. Basierend auf MRT- oder CT-Bildgebung kann die genaue Form des Knorpeldefekts erfasst und ein passgenau angepasstes Implantat gedruckt werden. Die Glykopolymer-Matrix dient dabei nicht nur als mechanisches Gerüst, sondern auch als bioaktives Umfeld, das die Chondrozyten in ihrer natürlichen Funktion unterstützt und die Integration mit dem umgebenden Gewebe fördert.

5 Anwendungsperspektiven und Marktpotenzial

Biofunktionelle Polymere adressieren mehrere Wachstumsmärkte mit erheblichem Innovationspotenzial:
Medizintechnik und regenerative Medizin: Der globale Markt für 3D-Bioprinting wächst mit über 20% CAGR. Personalisierte Knorpelimplantate, wirkstofffreisetzende ophthalmische Systeme und bioaktive Beschichtungen für Implantate haben hohes Marktpotenzial bei gleichzeitig verbesserter Patientenversorgung.
Pharmazeutische Formulierung: Die Stabilisierung von Protein-Therapeutika und mRNA-Impfstoffen ist eine Schlüsselherausforderung der modernen Pharmazie. Glykopolymer-basierte Formulierungsstrategien können die Haltbarkeit und Lagerstabilität empfindlicher Biomoleküle signifikant verbessern.
Funktionale Beschichtungen: Antimikrobielle, anti-adhesive oder drug-eluting Beschichtungen für medizinische Devices, Sensoren und technische Oberflächen. Der Markt für funktionale Beschichtungen im Gesundheitsbereich beträgt mehrere Milliarden Euro jährlich.
Intelligente Materialien: Sensormaterialien mit integrierter Erkennungsfunktion für Point-of-Care-Diagnostik, Umweltmonitoring oder Lebensmittelqualität. Die Integration biologischer Erkennungsmechanismen ermöglicht hochspezifische Sensoren ohne komplexe Elektronik.
Nachhaltige Hochleistungspolymere: Biofunktionalisierte, biobasierte Polymere mit kontrolliertem/verbessertem Abbauverhalten oder adaptiven Eigenschaften für Verpackung, Bauwesen und Mobilität.

6 Schlussfolgerungen und Ausblick

Biofunktionelle Polymere repräsentieren eine neue Materialklasse, die systematisch biologische Funktionsprinzipien mit synthetischer Polymerchemie vereint. Die Materialien gehen weit über klassische Bionik hinaus: Sie nutzen molekulare Erkennung, multivalente Wechselwirkungen und enzymatische Katalyse für völlig neue Materialeigenschaften, die mit konventionellen Ansätzen nicht erreichbar wären.
Die enge Kooperation zwischen angewandter Forschung am Fraunhofer IAP und universitärer Grundlagenforschung an der BTU Cottbus-Senftenberg schafft optimale Bedingungen für die Entwicklung solcher neuer Materialein.
Die Zukunft der Polymermaterialien liegt in der intelligenten Integration biologischer Funktionen. Biofunktionelle Polymere werden in den kommenden Jahren in immer mehr Anwendungsfeldern konventionelle Materialien ergänzen oder ersetzen, von der personalisierten Medizin über pharmazeutische Formulierungen und funktionale Oberflächen bis zu nachhaltigen Hochleistungspolymeren für die Kreislaufwirtschaft.

7 Literatur

[1]

DOI 10.5162/iCCC2026/5.3

Entwicklung von Plasmonischen On-Chip Sensoren
Development of Plasmonic On-Chip Sensors

Sebastian Reiter, Brandenburgische Technische Universität Cottbus-Senftenberg, Cottbus, reiter@b-tu.de
Akant Sengül, Brandenburgische Technische Universität Cottbus-Senftenberg, Cottbus, Germany, akant.senguel@b-tu.de
Paul-Gregor Nitsch, Brandenburgische Technische Universität Cottbus-Senftenberg, Cottbus, Germany, Paul-Gregor.Nitsch@b-tu.de
Fritz Berkmann, Brandenburgische Technische Universität Cottbus-Senftenberg, Cottbus, Germany, fritz.berkmann@b-tu.de
Carlos Alvarado Chavarin, IHP - Leibniz-Institut für innovative Mikroelektronik, Frankfurt Oder, Germany, alvarado@ihp-microelectronics.com
Jon Schlipf, IHP - Leibniz-Institut für innovative Mikroelektronik, Frankfurt Oder, Germany, schlipf@ihp-microelectronics.com
Christian Mai, IHP - Leibniz-Institut für innovative Mikroelektronik, Frankfurt Oder, Germany, cmai@ihp-microelectronics.com
Christian Wenger, IHP - Leibniz-Institut für innovative Mikroelektronik, Frankfurt Oder, Germany, wenger@ihp-microelectronics.com
Inga Anita Fischer, Brandenburgische Technische Universität, Cottbus, Germany, fischeri@b-tu.de

Kurzfassung

Diese Arbeit präsentiert die Entwicklung eines plasmonischen On-Chip Sensors, der als hochsensitiver Brechungsindexsensor für biosensorische Anwendungen konzipiert wurde. Der Sensor integriert ein plasmonisches Titannitrid (TiN) Nanolochgitter (NHA) mit einem Germanium (Ge) Photodetektor. Die optische Transmission durch das NHA wird überwiegend durch plasmonische Anregungen bestimmt und zeigt eine außergewöhnliche optische Transmission (EOT). Diese Resonanzen reagieren äußerst empfindlich auf Änderungen des Brechungsindexes der umgebenden Medien.

Abstract

This work presents the development of a plasmonic on-chip sensor as a highly sensitive refractive index sensor for biosensing applications. The device integrates a plasmonic titanium nitride (TiN) nanohole array (NHA) with a germanium (Ge) photodetector. The optical transmission through the NHA is predominantly determined by plasmonic excitations and exhibits extraordinary optical transmission (EOT). These resonances are highly sensitive to variations in the refractive index of the surrounding medium.

1 Introduction

There is a growing need for miniaturized on-chip optical sensors, which can be utilized, e.g., for biosensing in diverse fields such as agriculture, environmental diagnostics and healthcare. Here, we present recent results on a plasmonic sensor concept for applications in refractive index sensing. Our device is fabricated on a 200 mm Si platform, enabling both cost-effective realization and a high degree of reproducibility. We report on experimental sensor characterization and sensing performance, compare to simulation and discuss possible further applications of our devices.

2 Refractive Index Sensor

Optical properties of plasmonic resonances such as propagating surface plasmon polaritons at the interface between a metal and a dielectric or localized surface plasmon resonances in metallic nanoparticles sensitively depend on the refractive index of the surrounding dielectric medium. This has applications in refractive index sensing, however, commercialized sensors based on plasmonic excitations rely on bulky equipment such as microscopes or spectrometers for readout. Here, with the aim of miniaturizing this approach, we present an on-chip solution consisting of a plasmonic nanohole array as sensing element, whose optical properties change when exposed to dielectrics with different refractive indices, in combination with a Ge-photodetector that serves as a transducer [1] and converts the optical signal directly into a photocurrent. A schematic drawing of the device is presented in Figure 1.

DOI 10.5162/iCCC2026/5.3

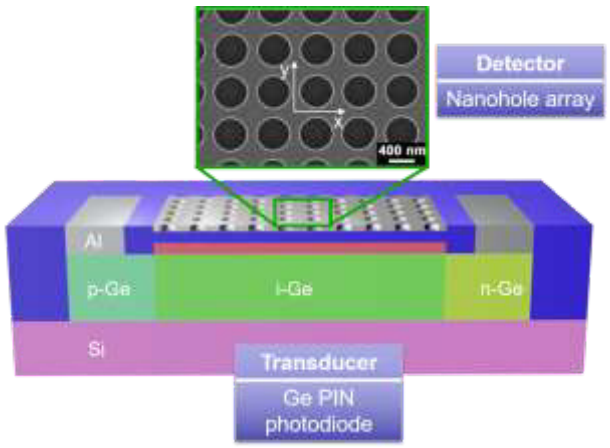


Figure 1 Schematic drawing of the concept of a plasmonic on-chip refractive index sensor consisting of a plasmonic TiN NHA and da Ge photodetector.

Compared to the realization of a proof-of-concept device [1], transitioning the device concept to large-scale production necessitates modifications for compatibility with CMOS manufacturing processes, introducing additional constraints in material choices, fabrication and design [2,3]. Our sensor can be fabricated using standard industrial complementary metal-oxide-semiconductor (CMOS) fabrication processes [2,3], enabling cost-efficient, high-yield production. In our devices, TiN as a CMOS-compatible metal is used for the fabrication instead of the more widely used Au and Ag.

2.1 Results and Discussion

Surface plasmon polaritons (SPPs) can be excited by illumination of the NHA, which enables momentum matching between the incident photons and the SPPs through the reciprocal lattice vectors of the periodic structure. This leads to EOT, which occurs as asymmetric resonance peaks known as Fano-resonances [3-5]. These resonances are highly sensitive to changes of the refractive index in the proximity of the NHA (Fig. 2).

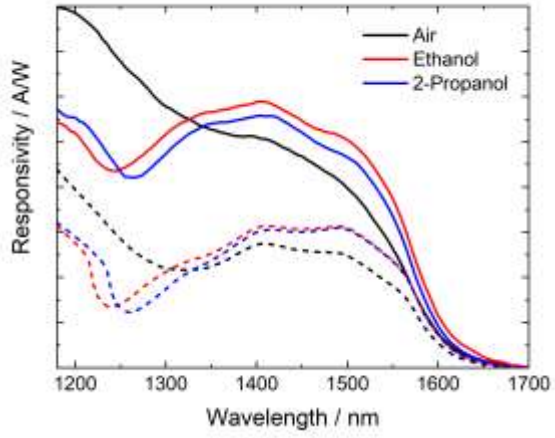


Figure 2 Responsivity spectrum of on-chip refractive index sensor immerged in ethanol (red) and 2-Propanol (blue).

In our experimental setup, shifts of the resonance dip are induced by altering the refractive index of the surrounding medium, notably, by immersing the sensor in ethanol (red) and 2-propanol (blue). The measured results (solid lines) compared to the simulated results (dashed lines) indicate a very high qualitative agreement and demonstrate the operability of the sensor concept. The difference between simulation and experiment results from the illumination conditions, which allow scattered light to enter the photodetector without passing through the nanohole array. The sensitivity to bulk refractive index changes, defined as the shift in resonance wavelength divided by the change in refractive index (RI), was measured to be ~900 nm/RIU in our sensors. This is confirmed by FDTD simulations.

3 Conclusion

Our results indicate that integrating NHAs with Ge photodetectors is an interesting approach for on-chip RI sensing. The integration of the NHA into a Ge-photodetector enables the use of the generated photocurrent as the output signal of the device. As a result of comparatively large damping in TiN, resonances in our NHAs are broad. Various strategies exist and are being implemented by us to improve the resonance shape while leaving the sensitivity unaffected. These comprise geometry changes (generation of quasi-BICs by changing the shape of the nanoholes) and material changes (deposition of a thin layer of Au in a post-CMOS deposition step). Furthermore, the presence of optical resonances in the responsivity spectra of our devices also have potential applications in on-chip spectrometers. Our work, thus, represents an important step in the development of cost-efficient on-chip sensors on the silicon platform.

4 Acknowledgements

This work was funded by Germany's Federal Ministry of Research, Technology and Space (BMFTR) under grant numbers: 16ES1128K, 16ME0420K, 16ES1131 16ME0424. The authors gratefully acknowledge the computing time granted through the North-German Supercomputing Alliance (HLRN) and provided on the supercomputing system. This work was supported by HLRN project ID bbp00044.

5 Literature

[1] L. Augel, Y. Kawaguchi, S. Bechler, R. Körner, J. Schulze, H. Uchida, and I. A. Fischer, "Integrated Collinear Refractive Index Sensor with Ge PIN Photodiodes," ACS Photonics **5**(11), 4586–4593 (2018).

[2] C. Mai, S. Marschmeyer, A. Peczek, A. Kroh, J. Jose, S. Reiter, I. Fischer, C. Wenger, and A. Mai, "Integration Aspects of Plasmonic TiN-based Nano-Hole-Arrays on Ge Photodetectorsin a

200mm Wafer CMOS Compatible Silicon Technology," ECS Transactions **109**, 35–46 (2022).

[3] C. Mai, A. Peczek, A. Kroh, J. Jose, S. Reiter, C. Wenger, and I. A. Fischer, "Towards a CMOS compatible refractive index sensor: cointegration of TiN nanohole arrays and Ge photodetectors in a 200 mm wafer silicon technology," Optics Express **32**(17), 29099–29111 (2024).

[4] S. Reiter, W. Han, C. Mai, D. Spirito, J. Jose, M. Zöllner, O. Fursenko, M. A. Schubert, I. Stemmler, C. Wenger, and I. A. Fischer, "Titanium Nitride Plasmonic Nanohole Arrays for CMOS-Compatible Integrated Refractive Index Sensing: Influence of Layer Thickness on Optical Properties," Plasmonics **18**(3), 831–843 (2023).

[5] S. Reiter, M. Ratzke, P.-G. Nitsch, C. Mai, D. Spirito, A. A. Corley-Wiciak, C. Wenger, and I. A. Fischer, "Optical Response of Titanium Nitride Plasmonic Nanohole Arrays: Impact of Square and Hexagonal Array Geometry, Pitch, and Nanohole Diameter," Plasmonics (2025).

DOI 10.5162/iCCC2026/5.4

Auf dem Weg zur radarbasierten Rinderüberwachung: Erste Überlegungen, Aufbauten und Ergebnisse
Towards Radar-Based Cattle Monitoring: Initial Considerations, Setups and Results

Markus Ulbricht, Vladica Sark, Rizwan Tariq Syed, Johann Paale Gose
IHP – Leibniz Institut für innovative Mikroelektronik,
Frankfurt (Oder), Deutschland,
[ulbricht | sark | syed | gose]@ihp-microelectronics.com
Gundula Hoffmann
ATB - Leibniz-Institut für Agrartechnik und Bioökonomie,
Postdam, Deutschland,
ghoffmann@atb-potsdam.de

Kurzfassung
Die Überwachung der Gesundheit von Nutztieren ist sowohl für die Agrarwirtschaft als auch aus ethischer Sicht unerlässlich. Automatisierte, sensorbasierte Verfahren können Tierärzte entlasten, eine lückenlose Kontrolle ermöglichen und Behandlungen präziser gestalten. Ein besonders vielversprechender Ansatz ist die Radarüberwachung, die ein kontaktloses, gleichzeitiges Monitoring von Herzfrequenz und Atmung mehrerer Tiere erlaubt. In dieser Arbeit setzen wir zwei ursprünglich für die Humanmedizin entwickelte Radarsysteme bei Kühen ein und präsentieren sowie diskutieren die erzielten Ergebnisse und mögliche Verbesserungen.

Abstract
Monitoring the health of livestock is essential both for the agricultural sector and from an ethical perspective. Automated, sensorbased approaches can ease the workload of veterinarians, enable continuous monitoring, and make treatments more targeted. A particularly promising method is radar-based monitoring, which allows simultaneous, contactless measurement of heart rate and respiration in multiple animals. In this work, we apply two radar systems originally developed for human medicine to dairy cows and present as well as discuss the obtained results and potential improvements.

1 Motivation und Einleitung

Deutschland ist nach wie vor stark von der Agrarwirtschaft geprägt. Im Jahr 2020 entfielen von den exportierten Waren im Gesamtwert von 1.205 Milliarden Euro rund 7 % auf Agrarprodukte. Tierische Erzeugnisse machten dabei etwa 61 % der Verkaufserlöse und rund 46 % des gesamten Produktionswertes der Landwirtschaft aus [1]. Angesichts der großen Zahl an Tieren in diesem Sektor können Krankheiten die Produktionsraten und die Qualität von Produkten wie Milch und Fleisch erheblich beeinträchtigen. Um solche Einflüsse zu vermeiden und zugleich das Wohl der Tiere zu gewährleisten, kommt der Überwachung der Tiergesundheit in der Nutztierhaltung eine zentrale Bedeutung zu. Aktuell findet die Überwachung im Wesentlichen durch den Landwirt selbst, sowie durch veterinärmedizinische Untersuchungen statt. Dies wird jedoch meist händisch an einzelnen Individuen durchgeführt und kann deshalb, aufgrund der Größe der Herde, nur stichprobenartig geschehen, woraus sich gewisse Unsicherheiten und damit potentielle Risiken ergeben. Dabei bietet die frühzeitige Warnung bei individuellen Abweichungen durchaus Vorteile, die sie eine schnelle Handlung und ggfs. Reduktion der Medikamentengaben ermöglicht. Darauf abzielende, automatisierte Methoden, wie z:B. RFID- oder im Pansen befindliche Bolus-Sensoren haben jedoch die wesentlichen Nachteile, dass sie, je nach Art, entweder invasiv, kostspielig, funktions- oder reichweitenbegrenzt sind. Radarbasierte Lösungen sind hier eine lohnenswerte Alternative. So versprechen aktuell untersuchte Lösungen mit auf FMCW (frequency modulated continuous wave) basierten MIMO (multiple input multiple output) Radaren die kontaktlose Überwachung der Vitalparameter mehrerer Tiere gleichzeitig, womit sie sich ideal für die Überwachung von Tieren in Herden eignet [2]. Das vorliegende Papier untersucht die Eignung zweier radarbasierter Lösungen für die Überwachung von Kühen: das in einem Vorläuferprojekt für die Humanmedizin entwickelte iCampus-Radar und ein kommerzielles und ebenfalls für Menschen entwickeltes Radar von Texas Instruments. Dabei präsentieren wir erste Versuchsaufbauten, zeigen vorläufige Messergebnisse und diskutieren geplante Schritte um die Qualität der bisherigen Messungen zu verbessern.

2 Stand der Technik

Aktuell werden in der Milchviehhaltung vor allem Beschelunigungs- und Temperatursensoren zur Überwachung der Tiergesundheit eingesetzt. Diese sind in

DOI 10.5162/iCCC2026/5.4

der Regel an einem Halsband oder per Pedometer am Bein der Tiere befestigt. Sie erfassen dabei vor allem das Tierverhalten und informieren den Landwirt, wenn es Abweichungen vom Normalverhalten (z. B. erhöhte Aktivität, reduzierte Wiederkauzeit) gibt. Neuere Entwicklungen ermöglichen jedoch auch die Überwachung von physiologischen Parametern am Einzeltier. So kann per Pansenbolus (z.B. smaXtec, Graz, Österreich) die Körpertemperatur und per Nasenclip (Gouna, Schönwalde) oder Infrarotkamera die Atemfrequenz gemessen werden. Die kontaktlose Überwachung von Nutztieren mittels Funkwellen findet zurzeit noch keinen breiten Einsatz, ist aber aufgrund ihrer wesentlichen Vorteile Gegenstand intensiver Forschungsarbeiten. Die Aktivitäten erstrecken sich dabei über verschiedene Technologien wie FMCW [3], UWB [4] oder WLAN [5] und zielen darauf ab, je nach Einsatzgebiet – etwa beim Menschen, bei Haus- oder Nutztieren – jeweils angepasste und optimale Ergebnisse zu erzielen. Bei Menschen liegt der Fokus häufig auf kritischen Anwendungsfällen, etwa im Krankenhaus oder im Fahrzeug [6], und teilweise existieren bereits erste kommerzielle Lösungen. Die Qualität der Ergebnisse wird dabei oft durch Verfahren wie Sensorfusion oder KI-basierte Methoden [6] erheblich verbessert. Für Tiere sind solche Systeme jedoch noch nicht etabliert, und es bestehen weiterhin mehrere komplexe Herausforderungen, die es zu lösen gilt.

3 Durchführung & Versuchstiere

Die für die vorliegenden Untersuchungen vorgenommenen Messungen an Kühen wurden in einem Milchviehstall der Lehr- und Versuchsanstalt für Tierzucht (LVAT) in Groß Kreutz durchgeführt. Bei dieser Anlage handelt sich um einen Liegeboxenlaufstall mit freiem Kuhverkehr und Spaltenboden. Der Stall beherbergt durchschnittlich 120 laktierende Milchkühe (1. bis 8. Laktation). Die Versuchstiere waren zwei Milchkühe der Rasse Holstein-Friesian: Kuh 138 (geb. 10.10.2019, 4. Laktation, 223. Laktationstag, durchschnittlich 40,1 kg Milch / Tag) und Kuh 251 (geb. 30.9.2021, 2. Laktation, 165. Laktationstag, 34,7 kg Milch / Tag).

4 iCampus-Radar

Im Rahmen des iCampus-Projekts wurde ein Dauerstrichradar (CW-Radar) zur Überwachung menschlicher Vitalfunktionen wie Atmung und Herzaktivität entwickelt. Dieses Radar erfasst minimale Hautverschiebungen, die durch Thoraxbewegungen und herzbedingte Vibrationen hervorgerufen werden. In diesem Beitrag präsentieren wir Testergebnisse, die mit demselben Radarsystem bei Tieren, insbesondere Kühen, erzielt wurden. Auch bei Kühen erzeugen Atmung und Herzaktivität messbare Oberflächenverschiebungen. Bewegungen entlang der radialen Achse des Radars werden besonders gut erfasst und können vom CW-Radar zuverlässig detektiert werden. Die radiale Verschiebung – d. h. die Verschiebung entlang der radialen Achse des Radars – kann wie folgt berechnet werden:

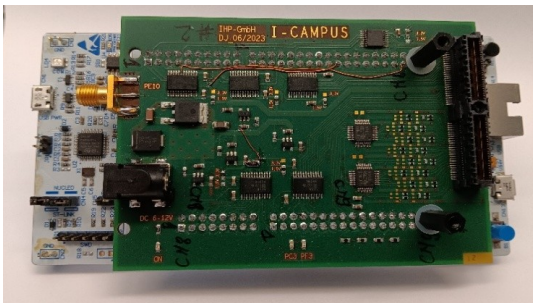


Bild 1: Datenerfassungsplatine

$$\Delta R = \frac{\lambda}{4\pi} \Delta\phi$$

Hierbei ist ΔR die geschätzte radiale Verschiebung des betrachteten Objekts, $\Delta\phi$ die gemessene Phasenverschiebung des empfangenen Radarsignals und λ dessen Wellenlänge. Die Phase des empfangenen Signals ist direkt proportional zur Entfernung zwischen Radar und Ziel. Das CW-Radar-Rücksignal umfasst die durch Atmung und Herzaktivität hervorgerufene Bewegung der Hautoberfläche des Tieres sowie Bewegungen, die durch allgemeine Körperbewegungen entstehen. Die Trennung von Vitalsignalen von Signalen, die durch die grobmotorischen Bewegungen des Tieres verursacht werden, ist eine anspruchsvolle Aufgabe. Es wurden verschiedene Methoden vorgeschlagen, darunter die empirische Modenzerlegung, Filtertechniken (z. B. Bandpass- und Kalman-Filter) und maschinelle Lernverfahren. Die Bandpassfilterung ist aufgrund ihrer Einfachheit und Eignung für die Implementierung auf eingebetteter Hardware oft die Methode der Wahl. Beim Bandpassfilterverfahren werden typischerweise zwei Filter entworfen und auf das empfangene Signal angewendet. Der Filter mit der niedrigeren Mittenfrequenz dient der Extraktion des Atmungssignals, während der Filter mit der höheren Mittenfrequenz die Herzaktivität isoliert. Bei dieser Methode ist es wichtig, dass sich Atmungs- und Herzfrequenz nicht überlappen. In Fällen, in denen Überlappungen auftreten können, können zusätzliche Signalverarbeitungstechniken eingesetzt werden, um die Komponenten effektiver zu trennen. Zum Zweck der Überwachung von Vitalparametern wurde ein „Single Input Multiple Output“ (SIMO) (1 TX und 8 RX-Antennen) CW-Radar [7] zusammen mit einem entwickelten Radardaten¬erfassungssystem verwendet. Das entwickelte Datenerfassungssystem Bild 1 verfügt insgesamt über 16 analoge Kanäle (bzw. 8 differentielle) zur Erfassung des analogen Radarsignals sowie über 32 General-Purpose-Input/Output-Kanäle zur Steuerung der Radar¬funktionen. Die maximale Abtastrate beträgt 32 kHz pro Kanal. Dieses Datenerfassungssystem wurde in Kombination mit dem oben ge-

DOI 10.5162/iCCC2026/5.4

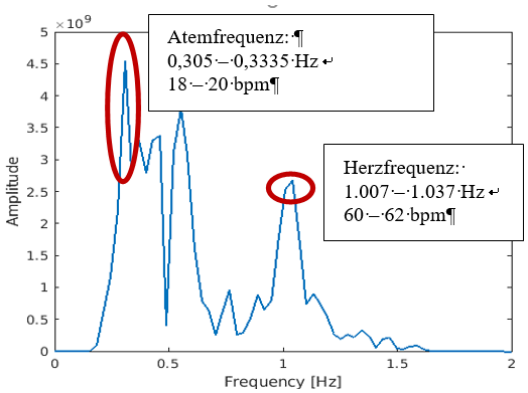


Bild 2: Spektrum des empfangenen Radarsignals

Bild 3: Aufbau zur Messung der Vitalparameter mit dem TI-AWR1642BOOST-Evaluierungsboard

nannten Radar für die ersten Messungen, d. h. die Datenerfassung, eingesetzt. Die ersten Messungen wurden am ATB Potsdam durchgeführt. Der Referenzwert für die Atemfrequenz wurde mithilfe eines Nasenclips (Gouna GmbH, Schönwalde, Deutschland) ermittelt, und der Referenzwert für die Herzfrequenz wurde mithilfe eines Faros 180 (Mega Electronics, Kuopio, Finnland). Die Referenz-Atemfrequenz lag im Bereich von 16 bis 20 Atemzügen pro Minute (bpm) und die Herzfrequenz im Bereich von 58 bis 62 Schlägen pro Minute (bpm). Die vorläufigen Ergebnisse sind in Bild 2 dargestellt. Die Abbildung zeigt das erhaltene Spektrum der Phase des Radarsignals. Es sind deutliche Peaks erkennbar – einer entspricht der Atemfrequenz und ein weiterer der Herzfrequenz. Aufgrund der geringen Frequenzauflösung wird angenommen, dass die tatsächliche Atem- bzw. Herzfrequenz zwischen den beiden höchsten Punkten des jeweiligen Peaks liegt. Die geschätzten Werte für die Atemfrequenz liegen zwischen 18 und 20 bpm und für die Herzfrequenz zwischen 60 und 62 bpm, was hervorragend mit den gemessenen Referenzwerten übereinstimmt.

5 TI Radar

Das Evaluierungsboard AWR1642BOOST von Texas Instruments (TI), das aus einem Ein-Chip-FMCW-Radarsensor besteht, wird zur Messung der Vitalparameter von Kühen verwendet. Radargeräte können winzige Brustbewegungen, die durch Atmung und Herzschlag verursacht werden, aus der Ferne erkennen, sodass keine physischen Sensoren oder tragbaren Geräte erforderlich sind. Durch die Analyse der Phasenänderungen in den reflektierten Radarsignalen kann das AWR1642BOOST die Atemfrequenz und Herzfrequenz in Entfernungen von wenigen Zentimetern bis zu mehreren Metern ermitteln.

5.1 Firmware und Messungen

Für erste Tests haben wir die von Texas Instruments (TI) bereitgestellte Firmware für Vitalparameter verwendet. Die Firmware enthält eine Demo-Binärdatei für die Messung von Vitalparametern und eine grafische Benutzeroberfläche zur Überwachung der entsprechenden Parameter. Die Binärdatei wird auf die Radarkarte AWR1642BOOST geflasht, wo die Rohdaten des Radars auf dem Chip verarbeitet werden. Die Firmware berechnet die Herzfrequenz und Atemfrequenz mithilfe verschiedener Signalverarbeitungsmethoden und überträgt die Ergebnisse über eine UART-Schnittstelle an einen PC. Anschließend werden die gemessenen Herzfrequenz- (HR) und Atemfrequenzwerte (BR) über eine Qt-basierte GUI angezeigt. Die HR- und BR-Daten werden ebenfalls als Binärdatei gespeichert, die zur weiteren Visualisierung und Analyse in MATLAB exportiert werden kann. Die von TI bereitgestellte Firmware schätzt die HR und BR anhand von vier Methoden: 1) FFT-basiert, 2) Peak-Zählung, 3) Erweitertes Frequenzband (4 Hz) und 4) Autokorrelationsbasiert. In dieser Studie wird die Peak-Zählmethode zur Messung der Vitalparameter von Menschen und Kühen betrachtet. Hierbei handelt es sich um eine Zeitbereichsmethode, bei der die Anzahl der Peaks in der gefilterten Wellenform gezählt und die Rate auf der Grundlage eines Zeitfensters berechnet wird. Die Radarplatte wird in einem Abstand von etwa 1,5 Metern zur Kuh platziert, wie in Bild 3 dargestellt. Für die Messungen am Menschen wurden die Experimente in einem Raum durchgeführt, in dem sich die Testperson in einem Abstand von etwa 1,5 Metern zum Radar befand.

5.2 Messungen zu BR und HR

Wir haben Messungen sowohl an Kühen als auch an Menschen durchgeführt. Ein MightySat-Pulsoximeter wurde verwendet, um die Referenzmessungen (Goldstandard) für HR und BR zu erhalten. Bild 4 zeigt die BR- und HR-Messergebnisse für Menschen, während Bild 5 die BR- und HR-Messergebnisse für Kühe zeigt. Eine kontinuierliche goldene Referenz für Herzfrequenz und Atemfrequenz konnte nicht aufgezeichnet werden, da keine speziellen Sensoren verfügbar waren. Im Allgemeinen liegt die Herzfrequenz einer erwachse-

DOI 10.5162/iCCC2026/5.4

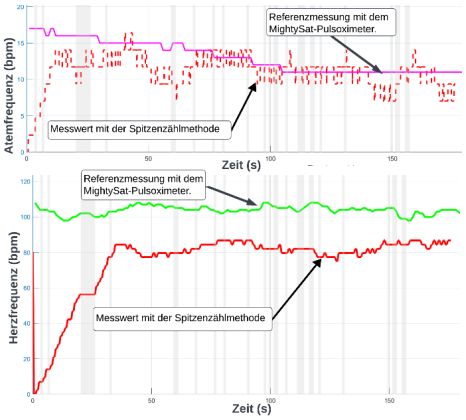


Bild 4: BR und HR Bestimmung beim Menschen

nen Kuh zwischen etwa 48 und 84 Schlägen pro Minute, und die Atemfrequenz liegt typischerweise zwischen 26 und 50 Atemzügen pro Minute, abhängig von der Rasse und den Umgebungsbedingungen. In unserem Aufbau wurde eine einzelne BR-Referenzmessung von 22 bpm mit einem Atmungssensor in der Nähe der Nasenlöcher der Kuh und eine einzelne HR-Referenzmessung von 62 bpm mit einem Stethoskop durchgeführt. Die schattierten Bereiche in Bild 4 und 5 zeigen Zeitintervalle, in denen sich der Mensch/die Kuh bewegte. Bild 4 zeigt die mit der Peak-Counting-Methode geschätzte menschliche BR. Die Peak-Counting-Messungen schwanken um den vom Mightysat-Gerät erhaltenen Referenzwert. Bei der menschlichen HR-Schätzung stabilisiert sich die Peak-Counting-Methode nach einiger Zeit, weist jedoch immer noch einen Fehler von mehr als 20 bpm auf. Im Fall von Kühen zeigt Bild 5 den mit der Peak-Counting-Methode geschätzten BR. Im Vergleich zur einmaligen Referenzmessung von 22 Schlägen pro Minute weist die Peak-Counting-Methode jedoch einen Fehler von etwa 7 Schlägen pro Minute auf. Bei der HR-Schätzung in Bild 5 stabilisiert sich die Peak-Counting-Methode nach einiger Zeit, weist jedoch immer noch einen Fehler von mehr als 10 Schlägen pro Minute auf.

6 Zusammenfassung und Ausblick

Die Untersuchungen zeigen, dass beide Radare grundsätzlich zur Erfassung von Vitalparametern bei Kühen geeignet sind. Das iCampus-Radar liefert dabei die zuverlässigsten und mit den Referenzwerten übereinstimmenden Ergebnisse. Das TI-Radar ist mit der vorhandenen Firmware nur eingeschränkt nutzbar, ermöglicht jedoch durch den Zugriff auf Rohdaten weitere Verbesserungen. Daher werden wir beide Ansätze weiterverfolgen – das iCampus-Radar zur präzisen Referenz und das TI-Radar als flexible Plattform für zukünftige Signalverarbeitung.

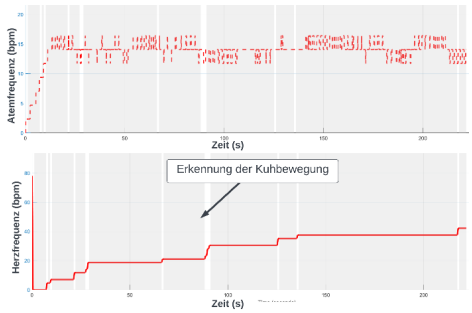


Bild 5: BR und HR Bestimmung beim Kühen

Literatur

[1] BMLEH, ‘Daten und Fakten - Land-, Forst- und Ernährungswirtschaft mit Fischerei und Wein- und Gartenbau’, Bundesministerium für Landwirt-schaft, Ernährung und Heimat, Jul. 2022. Ac-cessed: Nov. 18, 2025. [Online]. Available: https://www.bmleh.de/SharedDocs/Downloads/DE/Broschueren/daten-fakten-2022.pdf?__blob=publicationFile&v=8

[2] Ahmad, J. C. Roh, D. Wang, and A. Dubey, ‘Vi-tal signs monitoring of multiple people using a FMCW millimeter-wave sensor’, in 2018 IEEE Radar Conference (RadarConf18), Apr. 2018, pp. 1450–1455. doi: 10.1109/RADAR.2018.8378778.

[3] Y. Wang, Z. Wang, J. A. Zhang, H. Zhang, and M. Xu, ‘Vital Sign Monitoring in Dynamic Environ-ment via mmWave Radar and Camera Fusion’, IEEE Trans. on Mobile Comput., vol. 23, no. 5, pp. 4163–4180, May 2024, doi: 10.1109/TMC.2023.3288850.

[4] P. Wang et al., ‘Non-Contact Vital Signs Monitoring of Dog and Cat Using a UWB Radar’, Animals, vol. 10, no. 2, p. 205, Jan. 2020, doi: 10.3390/ani10020205.

[5] J. Liu, Y. Wang, Y. Chen, J. Yang, X. Chen, and J. Cheng, Tracking Vital Signs During Sleep Lever-aging Off-the-shelf WiFi. 2015, p. 276. doi: 10.1145/2746285.2746303.

[6] A. El Abbaoui, D. Sodoyer, and F. Elbahhar, ‘Contactless Heart and Respiration Rates Estimation and Classification of Driver Physiological States Using CW Radar and Temporal Neural Networks’, Sensors, vol. 23, no. 23, p. 9457, Nov. 2023, doi: 10.3390/s23239457.

[7] B. Sütbas, H.J. Ng, M.H. Eissa, C. Carta, G. Kahmen, ‘Vital Signs Monitoring using 61.4 GHz CW MIMO Radar Sensor with LO-Scalable Low-Voltage Low-Power SiGe BiCMOS Chipset’, Proc. 54th European Microwave Conference (EuMC 2024), 497 (2024)

DOI 10.5162/iCCC2026/5.5

CHEMOFLUIDISCHE SCHALTKREISE – PLATTFORM ZUR AUTOMATISIERUNG VON MEHRSTUFENPROTOKOLLEN ZUR DIAGNOSE VON ATEMWEGSINFEKTIONEN

CHEMOFLUIDIC CIRCUITS – A PLATFORM FOR AUTOMATING MULTISTEP PROTOCOLS FOR THE DIAGNOSIS OF RESPIRATORY INFECTIONS

Andreas Richter[1,3], Mario Menschikowski[2,3], Denise Gruner[1,] Mohammed Shahadha[1,3], Markus Friedemann[2,3], Franziska Obst[1,3], Theo Spitzner[1,3], Johanna Dohndorf[1], Johannes Paul[1,3]
[1]Dresden University of Technology, Chair of Microsystems, Germany, andreas.richter7@tu-dresden.de
[2]University Hospital Carl Gustav Carus Dresden, Germany, mario.menschikowski@ukdd.de
[3]Diamics GmbH, Dresden, Germany, Johannes.paul@diamics.com

Kurzfassung

Die COVID-19-Pandemie hat den Bedarf an skalierbaren, dezentralen Point-of-Care-Diagnostikplattformen verdeutlicht. Chemofluidik bietet hierfür ein neuartiges Konzept: Stimuli-responsive Polymere fungieren als chemofluidische Transistoren und ermöglichen die autonome Steuerung mikrofluidischer Prozesse ohne elektronische Off-Chip-Kontrolle. Auf Basis mehrlagiger Integrations- und Drucktechnologien lassen sich komplexe logische und analoge Schaltungen realisieren, mit denen man vollständig integrierte POCT inklusive Probenaufbereitung, isotherme Amplifikation und Multiplex-Detektion auf einem Chip automatisieren kann. Die chemofluidische Hardware-Zeitsteuerung gewährleistet ein robustes, reproduzierbares Ablaufprogramm und zeigt das Potenzial großflächig integrierter, autonom arbeitender Diagnostiksysteme.

Abstract

The COVID-19 pandemic has highlighted the urgent need for scalable, decentralized point-of-care diagnostic platforms. Chemofluidics offers a novel approach in this context: stimuli-responsive polymers act as chemofluidic transistors, enabling autonomous control of microfluidic processes without electronic off-chip regulation. Based on multilayer integration and printing technologies, complex logical and analog circuits can be implemented to automate fully integrated POCT systems, including sample preparation, isothermal amplification, and multiplex detection on a single chip. The intrinsic chemofluidic hardware timing ensures a robust and reproducible workflow, demonstrating the potential of large-scale integrated, self-contained diagnostic systems.

1 Einleitung

Die COVID-19-Pandemie hat verdeutlicht, dass die molekulare Diagnostik ein essenzielles Instrument für das präzise Testen, Überwachen und Eindämmen pandemischer wie epidemischer Ereignisse darstellt [1]. Gleichzeitig zeigte sich jedoch, dass selbst hervorragend ausgestattete Zentrallaboratorien in Industrieländern aufgrund begrenzter Skalierbarkeit, logistischer Abhängigkeiten und komplexer Prozessketten rasch an ihre Kapazitätsgrenzen stoßen können. Diese strukturellen Engpässe führten zu Verzögerungen in der Testauswertung und machten die Verwundbarkeit zentralisierter Diagnostikstrukturen deutlich. Vor diesem Hintergrund ist die Entwicklung dezentraler Point-of-Care-Plattformen, die idealerweise den ASSURED-Kriterien der World Health Organization entsprechen, zu einem prioritären Ziel der Lab-on-Chip-Community für die kommende Dekade geworden [2]. Solche tragbaren Point-of-Care-Tests (POCT) sollen erschwinglich, sensitiv, spezifisch, benutzerfreundlich, schnell, robust, geräteunabhängig und einfach distribuiert sein, sodass sie prinzipiell von jedermann und nahezu überall eingesetzt werden können. Dadurch ließe sich der Flaschenhals zentralisierter Laborinfrastruktur konzeptionell umgehen und zugleich eine zuverlässige Diagnostik in ressourcenlimitierten Umgebungen ermöglichen.

Trotz intensiver Forschungsaktivitäten existieren bislang nur wenige Plattformen, die tatsächlich portable, feldtaugliche und weitgehend autonom arbeitende Geräte für die molekulare Infektionsdiagnostik bereitstellen [3]. Die Arbeitsgruppe von N. Pamme demonstrierte beispielsweise ein Lab-on-a-Chip-System auf Basis magnetischer Beads mit integrierter RNA-Extraktion und Amplifikation mittels RT-LAMP, das den Nachweis von 470 Kopien genomischer RNA innerhalb einer Stunde ermöglichte [4]. Hinsichtlich Nachweischemie und analytischer Sensitivität nicht unmittelbar vergleichbar ist das vollständig durch Kapillarschaltungen automatisierte SARS-CoV-2-Antikörperassay mit enzymatischer Signalverstärkung auf ELISA-Basis der Gruppe von D. Juncker, das ebenfalls innerhalb von etwa einer Stunde Ergebnisse liefert [5]. Beide Ansätze verdeutlichen die Fortschritte auf dem Gebiet integrierter mikrofluidischer Diagnostik, zeigen jedoch zugleich die anhaltende Abhängigkeit vieler Systeme von externer Instrumentierung oder elektronischer Steuerung.

DOI 10.5162/iCCC2026/5.5

Eine vielversprechende zusätzliche Alternative zur elektroniklosen Automatisierung komplexer mikrofluidischer Prozesse stellt die Chemofluidik dar, die im Folgenden als eigenständiges Konzept vorgestellt und diskutiert wird.

2 Chemofluidische Schaltkreise

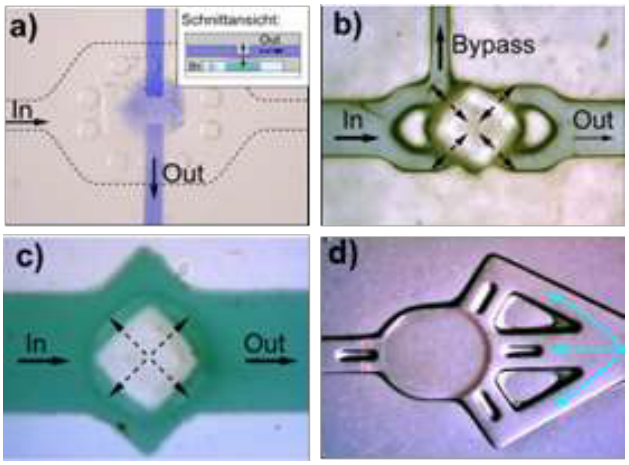


Bild 1 Chemofluidische Bauelemente. a) Membranisolierter Volumenphasenübergangstransistor, b) Volumenphasenüberganstransistor, c) chemofluidisches Schließerventil, d) chemofluidische Diode nach [6-8,18]

Die Chemofluidik als mikrofluidische Lab-on-Chip-Plattform besitzt eine konzeptionelle Alleinstellung, da sie bislang das einzige Schaltkreiskonzept darstellt, das fluidische Prozesse durch unmittelbares Feedback zwischen Bauelement und Prozessmedium steuert und dabei vollständig ohne eine elektronische Off-Chip-Steuerung auskommt. Im Zentrum dieses Ansatzes stehen stimuli-sensitive Polymere – insbesondere Hydrogele wie PNIPAAm-basierte Systeme sowie lösliche Polymere wie Polyethylenglykol (PEG) – die eine ausgeprägte, reversible Volumen- oder Eigenschaftsänderung in Abhängigkeit von chemischen Stimuli wie Lösungsmittelkonzentration, pH-Wert oder Salzgehalt zeigen. Diese intrinsisch aktiven Materialien fungieren zugleich als Sensor und Aktor und ermöglichen so eine direkte Kopplung zwischen chemischer Informationsdomäne und fluidischer Transportdomäne. Der chemische Zustand des Prozessmediums wird unmittelbar in eine mechanische Antwort – etwa das Öffnen oder Schließen eines Strömungspfades – übersetzt, wodurch chemische Information auf Chipebene verarbeitet werden kann.

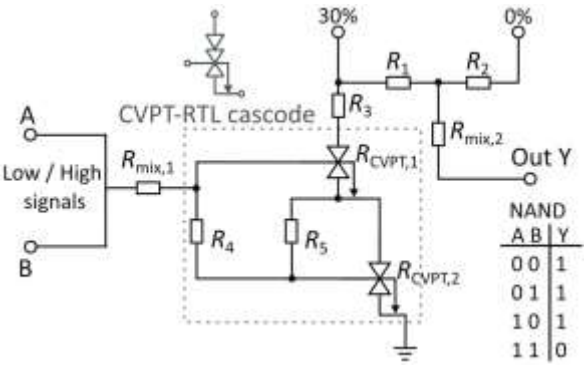


Bild 2 Schaltplan einer chemofluidischen Kaskode auf Basis von chemischen Volumenphasenübergangstransistoren nach [8]

Die daraus abgeleitete Bauelementepalette ist außerordentlich breit und umfasst unter anderem hydrogelbasierte Pumpen, Chemostatventile mit einstellbaren Schwellenwerten [9], chemofluidische Transistoren [10], Dioden, abstimmbare Mikrofilter und -poren [11,12], Speicherelemente sowie Analog-Digital-Wandler [13] (siehe auch **Bild 1**). Insbesondere der chemofluidische Transistor – realisiert als Volumenphasenübergangstransistor oder als membranisolierter Volumenphasenübergangstransistor (MIS-CVPT) – bildet das funktionale Pendant zum elektronischen Transistor und stellt die zentrale skalierbare Komponente einer integrierten chemofluidischen Schaltungstechnik dar. Analog zur Mikroelektronik lassen sich durch geeignete Grundschaltungen Hardwareinstruktionen implementieren; durch Kaskadierung und modulare Verschaltung entstehen komplexe Programmsequenzen mit autonomer Signalpropagation im chemischen und fluidischen Bereich.

Technologisch basiert die Chemofluidik auf einer Mehrlagentechnologie, die strukturell an mikroelektronische Integrationsverfahren erinnert. Fluidische Kanäle, Membranen, Aktorschichten und funktionale Polymerstrukturen werden in übereinanderliegenden Ebenen präzise ausgerichtet und miteinander gekoppelt. Die Integration aktiver Polymerbauelemente kann dabei sowohl photolithografisch – beispielsweise durch *in-situ*-Polymerisation strukturierter Hydrogele in definierten Kammern – als auch mittels additiver Multimaterialdruckverfahren erfolgen. Insbesondere der Multimaterialdruck erlaubt die flexible, maskenlose Herstellung funktionaler Schaltungen mit verschiedenen aktiven Materialien auf großflächigen Substraten und eignet sich damit besonders für die schnelle Prototypisierung und Kleinserienfertigung [14]. Photolithografische Verfahren hingegen bieten hohe laterale Präzision und Reproduzierbarkeit und ermöglichen die Realisierung komplexer, dicht integrierter Schaltungen mit vergleichbaren Toleranzen wie in der Halbleitertechnik [15]. Beide Ansätze eröffnen perspektivisch eine skalierbare Schaltkreisfertigung im Sinne einer chemofluidischen Integrationstechnologie.

Die chemofluidische Schaltungstechnik profitiert dabei in erheblichem Maße von Konzepten der elektronischen

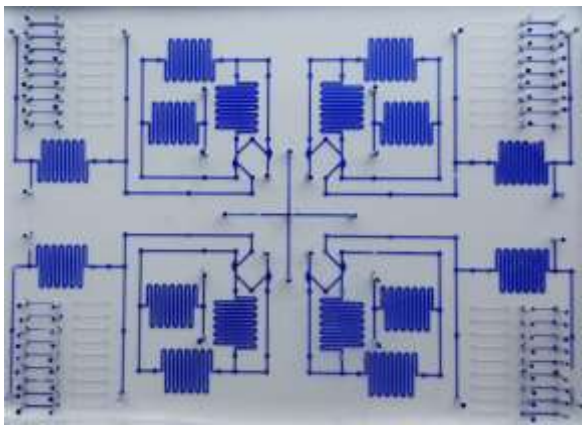


Bild 3 Chemofluidischer Chip mit 4 Kaskoden nach [8]

DOI 10.5162/iCCC2026/5.5

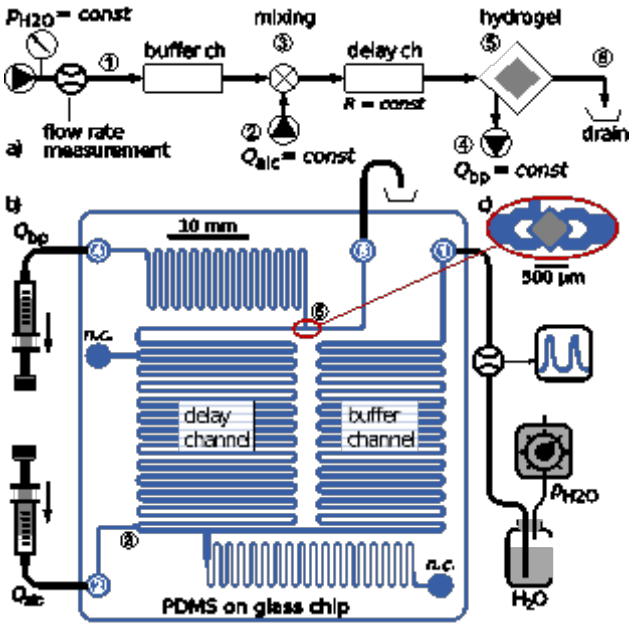


Bild 4 Chemofluidischer Verzögerungsleitungsoszillator nach [16]

Schaltungstheorie. So existieren zwei grundlegende Transistorkonzepte: Der membranisolierte Volumenphasenübergangstransistor weist funktionale Analogien zum Feldeffekttransistor (FET) auf, da ein chemisches Steuersignal in einem Kontrollkanal den Strömungswiderstand zwischen Source und Drain moduliert. Der nicht-membranisolierte Volumenphasenübergangstransistor realisiert hingegen eine direktere Kopplung zwischen chemischer und fluidischer Domäne. In diskreter Schaltungstechnik konnte auf dieser Basis eine resistor-transistor-inspirierte Logik aufgebaut werden, die eine umfangreiche Palette logischer Gatter (INV, AND, OR, NAND, NOR), sequenzielle Logik mit verschiedenen Flipflops sowie autonome Oszillatoren (**Bild 4**) umfasst [7, 8]. Darüber hinaus wurden analoge Schaltungen wie Mischer, Verstärker, Komparatoren und Kaskoden (**Bilder 2** und **3**) demonstriert, wodurch sich die chemofluidische Plattform nicht nur für digitale, sondern auch für kontinuierliche Signalverarbeitung eignet.

Der Übergang von einzelnen Bauelementen zu hochintegrierten chemofluidischen Schaltungen bleibt jedoch eine erhebliche Herausforderung. Ursache hierfür sind ausgeprägte Materialnichtlinearitäten, diffusionsgetriebene Dynamiken sowie komplexe domänenübergreifende Wechselwirkungen auf materialwissenschaftlicher, physikalischer, chemischer und teilweise biologischer Ebene. Die starke Kopplung von Strömungsmechanik, Transportprozessen und nichtlinearer Polymerthermodynamik führt zu dynamischen Effekten wie Hysterese, Schwellenverhalten oder Bifurkationen, die eine rein empirische Auslegung komplexer Schaltungen erschweren. In mehreren Arbeiten erwiesen sich daher ausschließlich experimentelle Optimierungsstrategien als unzureichend, was die frühzeitige Entwicklung domänenübergreifender Simulations- und Entwurfswerkzeuge motivierte. Hierzu zählen netzwerkbasierte Modellierungsansätze in Analogie zur elektrischen Schaltungssimulation ebenso wie bifurkationsbasierte Analysen zur Identifizierung stabiler Arbeitspunkte und Oszillationsregime [17]. Erst durch die Kombination aus skalierbarer Mehrlagenfertigung, präziser Integration aktiver Materialien und rechnergestütztem Schaltungsentwurf wird die Perspektive einer großflächig integrierten, autonomen chemofluidischen Schaltungstechnik realistisch [18].

3 Chemofluidischer POCT

Die derzeitigen Goldstandard-Nukleinsäuretests erfordern eine Vielzahl separater Prozessschritte, die durch geschultes Personal an spezifischer Instrumentierung in Laborumgebungen durchgeführt werden müssen. Dazu zählen Probenlyse, Nukleinsäurebindung und -reinigung, Elution, Reagenzvorbereitung, Amplifikation und Detektion – meist verteilt auf mehrere Geräteplattformen. Diese fragmentierte Prozesskette ist zeit- und ressourcenintensiv und limitiert insbesondere bei hoher Testnachfrage die Skalierbarkeit der Diagnostik. Unser mikrofluidischer Testchip integriert und automatisiert diese Schritte vollständig auf einer einzigen Plattform.

Nach Einbringen einer geeigneten Patientenprobe erfolgt zunächst die Freisetzung der Zielnukleinsäuren mittels eines geeigneten Lyse- und Bindesystems sowie deren Anbindung an eine integrierte Festphase innerhalb des Chips. In einem nachgeschalteten Reinigungsschritt werden gebundene Nukleinsäuren durch definierte Waschbedingungen von störenden Matrixbestandteilen und potenziellen Inhibitoren befreit. Anschließend erfolgt die Elution der aufgereinigten Nukleinsäure und deren Überführung in die nachfolgende Prozessstufe. Dort wird das Eluat mit einem isothermen Amplifikationssystem kombiniert, in separate Reaktionskammern dosiert, diese verschlossen und unter konstanten Temperaturbedingungen amplifiziert. Die Plattform ist grundsätzlich für die parallele Detektion mehrerer respiratorischer Zielsequenzen ausgelegt und erlaubt eine zeitnahe Auswertung mit einer analytischen Leistungsfähigkeit, die mit etablierten laborbasierten Referenzmethoden vergleichbar ist.

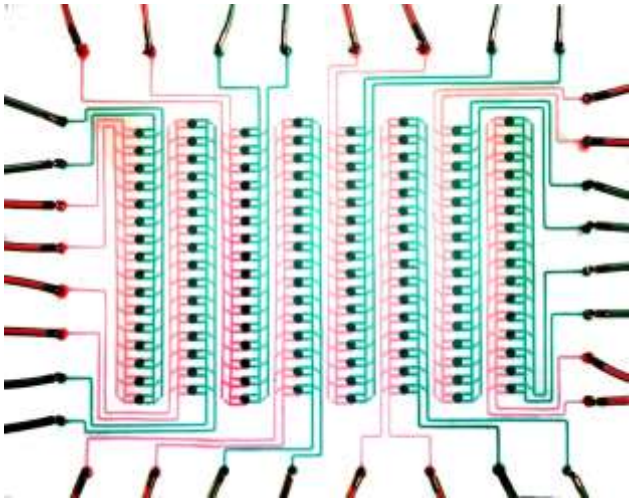


Bild 5 Chemofluidischer Schaltkreis mit 2.096 Schaltelementen und 384 Reaktionskaskaden für äquidistante Untersuchungen von Bioreaktoren nach [18]

Die gesamte Fluidhandhabung der durch eine Druckquelle angetriebenen Flüssigkeiten wird von einem Netzwerk chemofluidischer Schalter [6, 18] gesteuert, das auf stimuli-responsiven Materialien wie Hydrogelen und löslichen Polymeren basiert. Diese intrinsisch aktiven Materialien koppeln chemische Zustandsgrößen direkt an mechanische Schaltvorgänge und ermöglichen dadurch eine autonome Prozessführung ohne elektronische Off-Chip-Steuerung. Die zugrunde liegenden Basisschaltungen [7,8] übernehmen das zeitlich definierte Zu- und Abschalten einzelner Flüssigkeitsströme, das sequentielle Mischen und Dosieren von Reagenzien sowie deren gerichtete Weiterleitung durch das mikrofluidische Netzwerk. Darüber hinaus realisieren sie das zuverlässige Verschließen von Reaktionskammern und Schnittstellen, sodass kontrollierte Reaktionsbedingungen für die Amplifikation gewährleistet sind.

Der zeitliche Ablauf des gesamten Chipprotokolls wird durch die Dynamik der chemofluidischen Schalter und die definierten Strömungsparameter innerhalb des Systems bestimmt. Die materialintrinsische Schaltkinetik fungiert dabei als integrierte „Hardware-Zeitsteuerung“, wodurch ein reproduzierbares und robustes Prozessprogramm entsteht. Das implementierte chemofluidische Ablaufkonzept zeigt eine hohe Stabilität gegenüber moderaten Schwankungen von Betriebsparametern wie Druck, Temperatur oder Probenzusammensetzung.

Obwohl die Chips aus einer Vielzahl funktional differenzierter Lagen aufgebaut sind, lassen sie sich kosteneffizient durch eine Kombination aus Filmstrukturierungs-, Druck- und lithografischen Prozessen herstellen [14]. Diese Mehrlagentechnologie erlaubt die Integration fluidischer Kanäle, funktionaler Membranen, Dichtstrukturen und aktiver Polymerbauelemente in einem skalierbaren Fertigungsprozess und bildet damit die Grundlage für eine wirtschaftliche Realisierung hochintegrierter, autonom arbeitender mikrofluidischer Diagnostikplattformen.

4 Literatur

[1] Baldeh, M., *et al*.,”Lessons from the pandemic: new best practices in selecting molecular diagnostics for point-of-care testing of infectious diseases in sub-Saharan Africa”, *Expert Review of Molecular Diagnostics*, 2024, **24** (3), 153.

[2] N. Pamme, “Microfluidics for clinical diagnostics and environmental analysis in resource-limited settings”, MicroTAS 2023, October 13-17, 2023, Katowice, Poland.

[3] Moetlhoa, B., *et al*., “REASSURED diagnostics at point-of-care in sub-Saharan Africa: A scoping review”, *PLOS Glob Public Health*, 2023, **3** (6), e0001443.

[4] Rodrigues-Mateos, P., *et al*., “A lab-on-a-chip platform for integrated extraction and detection of SARS-CoV-2 RNA in resource-limited settings”, *Analytica Chimica Acta*, 2021, **1177**, 338758.

[5] Yafia, M., *et al*., “Microfluidic chain reaction of structurally programmed capillary flow events”, *Nature*, 2022, **605**, 464.

[6] Beck, A. *et al*., "Fundamentals of Hydrogel-Based Valves and Chemofluidic Transistors for Lab-on-a-Chip Technology: A Tutorial Review," *Adv. Mater. Technol*., 2023, **8**, 2200417

[7] Frank, P., *et al*., "Autonomous Integrated Microfluidic Circuits for Chip-Level Flow Control Utilizing Chemofluidic Transistors," *Adv. Funct. Mater*., 2017, 27, 1700430

[8] Beck, A. *et al*., "Logic Circuits Based on Chemical Volume Phase Transition Transistors for Planar Microfluidics and Lab-on-a-Chip Automation," *Adv. Mater. Technol*., 2022, **7**, 2200185

[9] Richter, A. *et al*., “Controlled Double-Sensitivity of Microgels Applied to Electronically Adjustable Chemostats”, *Adv. Mater*., 2007, **19**, 1109–1112

[10] Frank, P. *et al*., Integrated Microfluidic Membrane Transistor Utilizing Chemical Information for On-Chip Flow Control, *PLoS One*, 2016, **11**, 1–17

[11] Ehrenhofer, A. *et al.*, “Permeation control in hydrogel-layered patterned PET membranes with defined switchable pore geometry - Experiments and numerical simulation,” *Sens. Actuat. B*, 2016, 232, 499–505

[12] Haefner, S. *et al*., “Chemically Controlled Micro-Pores and Nano-Filters for Separation Tasks in 2D and 3D Microfluidic Systems”, *RSC Advances*, 2017, **7**, 49279 – 49289

[13] Haefner, S. *et al*., “Smart hydrogels as microfluidic storage elements with sampling readout functionality in discontinuous microfluidic systems”, *Lab Chip*, 2016, **16**, 3977-3989

[14] Kutscher, A., *et al*., "Fabrication of Chemofluidic Integrated Circuits by Multi-Material Printing", *Micromachines*, 2023, **14**, 699

[15] Haefner, S. *et al*., “High integration of microfluidic circuits based on hydrogel microvalves for MEMS control”, *Adv. Mater. Technol*. 2017, 1700108

[16] Paschew, G. *et al*., ”Autonomous chemical oscillator circuit based on bidirectional chemical-microfluidic coupling”, *Adv. Mater. Technol.*, 2016, **1**, 1600005

[17] Páez Chávez, J. *et al*., “A New Self-Excited Microfluidic Oscillator Based on Stimuli-Responsive Hydrogels: Mathematical Modeling and Dynamic Behavior”, *Appl. Math. Modelling*, 2016, **40**, 9719-9738

[18] Greiner, R. *et al*., “Fluidic microchemomechanical integrated circuits processing chemical information”, *Lab Chip*, 2012, **12**, 5034 - 5044

DOI 10.5162/iCCC2026/6.2

GigaCoupler – A scalable Opto-Isolator with Gbps-Data Transmission

Kai Schmieder, Tobias Schneider, René Kirrbach,
Fraunhofer Institute for Photonic Microsystems IPMS, Dresden, Germany
kai.schmieder@ipms.fraunhofer.de

Abstract

Optocouplers are widely used for signal transmission between galvanically separated circuit parts. For decades, these have been an integral component of electronic modules. The wide selection becomes limited, and costs rise disproportionately for data rates >50 Mbps or isolation voltage exceeds several kV. This is where the GigaCoupler comes in. It is a novel short-range optical transceiver technology. In this paper, we introduce the concept and describe the implementation. Results of initial testing show that a data rate of 1.25 Gbps is achievable with an isolation voltage in the double-digit kV-range and a footprint of only 0.5 cm². Potential applications such as optical probes, isolated A/D or D/A converters; an alternative to circuit boards with embedded waveguides and a rotary usage are in the focus of this contribution.

1 Introduction

Optocouplers transfer signals using light waves. The in- and output are electrically isolated. Typically, voltage differences of up to 5 kV are possible between the connecting legs [1]. Usually, uni-directional transmission is supported by off-the-shelf optocouplers. Typically, data rates of around 10 Mbps are already considered as high-speed [2]. The selection is very limited when a data rate above 50 Mbps is required [3].
High voltage isolation requires a large form factor, design restrictions and a limited data rate. Such a component is shown in **Figure 1**.

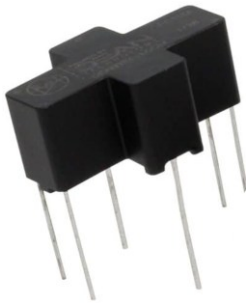


Figure 1 Specific optocoupler package [4]

If both are of interest: a high insulation and a high data rate, then “SFP” (Small Form Factor Pluggable) modules may be used. They consist of two optical transceiver-modules. In addition, these support bi-directional communication. They were developed for network connections and therefore suitable for low-loss data transmission over longer distances. A specific embodiment in this area is shown in **Figure 2**. It is called “SFP+ AOC (Active Optical Cable)” and has its main area of application for high-speed connections within racks and between neighboring racks [5]. The connection to a printed circuit board (PCB) is established by the required port cages. A sample can be seen in **Figure 3**. The required area for a single SFP module is 8 cm². However, for an opto-coupler application, two of them are required, and the bending radius of the fiber cable should not be neglected. Due to large form factors and high costs, SFP solutions are only a makeshift for intra- or inter-board connection.

Figure 2 SFP+ and AOC [5]

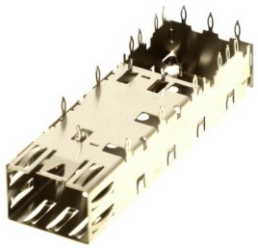


Figure 3 typical SFP cage [6]

Our solution represents a new approach which is called “GigaCoupler”. The transmitter (Tx) and the receiver (Rx) are spatially separated from each other. The GigaCoupler contains a second link for full-duplex functionality. The outward and the return path overlap spatially. This goes along with 2 benefits: it minimizes the required space, and the system becomes rotation capable. In contrast to the SFP solution, no optical fiber is required which reduces form factor, costs and complexity. All that is required is a line of sight between both transceivers. The GigaCoupler enables a surface-mountable assembly. **Figure 4** illustrates the use case of an intra-board link. The system is very robust, hence positioning tolerances are no challenge. Furthermore, the isolation can be tailored by setting the distance between the transceivers. Even with just a few centimeters,

DOI 10.5162/iCCC2026/6.2

the isolation voltage far exceeds that of conventional optocouplers. Leakage currents can be effectively kept small by inserting a slot like depicted in **Figure 4**.

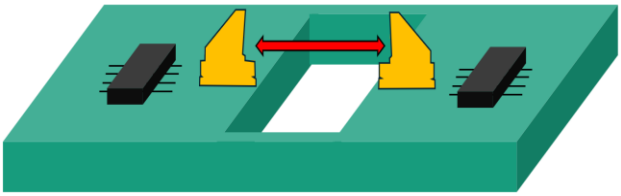


Figure 4 GigaCoupler integration scheme as intra-board link

The introduction of the GigaCoupler concept and proof of feasibility are part of a previous publication [7]. This article focuses on the analysis of potential applications.

2 Implementation of the GigaCoupler concept

The lens is a key component of the transceiver. It was carefully designed to meet the specific requirements of the GigaCoupler application. The lens has been shaped using an ultra-precision molding tool. An easy moldable polymer material was selected for the proof-of-concept phase. **Figure 5** shows the transceiver package consisting of free form lens and QFN (Quad flat no leads) substrate. That QFN-base enables surface mounting.

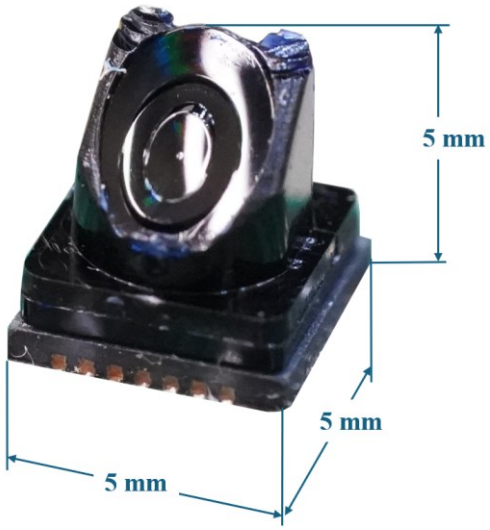


Figure 5 View of a GigaCoupler transceiver

The transceiver was soldered onto PCB. Two of these are placed in front of each other as shown in **Figure 6**. In order to assess the quality of the transmission link, the occurring bit errors are determined. The test also examines which misalignment between the transceiver-couple can be balanced by these before an increase in bit error rate (BER) is observed.
The data transfer was considered acceptable for a maintained error rate of max. 1 on a billion transferred bits. The acceptable error rate cannot be specified in general terms. This depends on the application scenario and here especially on the communication protocol and the possible error correction.

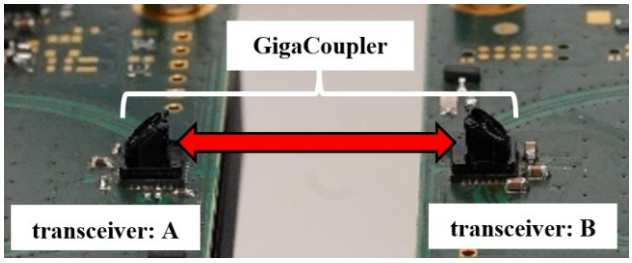


Figure 6 GigaCoupler integration as inter-board link (distance of transceivers: 23 mm)

3 Implementation types and potential application areas

There are many conceivable scenarios for the GigaCoupler. These range from feedback control in high-voltage power supplies, high speed digital in-/ output at peripheral interfaces of computers or isolated interfaces of PLCs (Programmable Logic Controller) to many other application fields. Following, a few areas of application are explained in detail to illustrate potential use cases of this product.

3.1 Isolated probing system

Probes are essential accessories for oscilloscopes, power analyzers, spectrum analyzers a similar equipment. Certain applications require a galvanic isolation of the DUT (Device under Test) from ground and oscilloscope. This may be the case for instance when modern semiconductor components like IGBTs (Insulated Gate Bipolar Transistor) with voltage levels up to 6.5 kV with stately shorter slew rates or motor drives should be characterized. These isolated probes are featured by a high common mode rejection ratio (CMRR). The elimination of electrical return paths gives the chance to keep this value as high as possible. High quality probes allow a CMRR value in the mid 5-digit range at a frequency of 1 GHz [8], enabling the measuring of small signals despite a high voltage and fast switching environment. The RT-ZISO from Rohde & Schwarz is such a high-end probing system [8]. The fiber architecture used for this purpose isolates the DUT from the measuring setup at a frequency of up to 1 GHz. **Figure 7** illustrates the structure of that probing system.

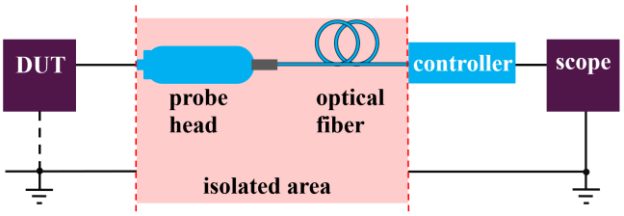


Figure 7 Isolating using optical fiber – illustration based on [8]

The substitution of optical fiber by GigaCoupler would lead to a more compact and inexpensive probe setup. The

DOI 10.5162/iCCC2026/6.2

arrangement of controller device and probe head inside of one common unit can save space which is shown in **Figure 8**.

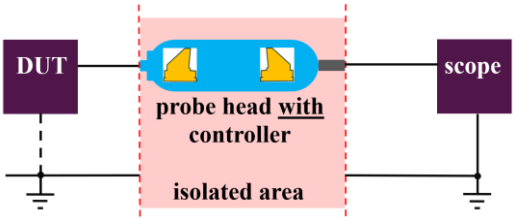


Figure 8 Probe head with integrated controller unit – isolated by GigaCoupler

3.2 Power analyzer

The precise determination and analysis of power shares is of great importance for the efficient generation of energy and its use. The measurement of active power is performed by the simultaneously measurement of current and voltage. The measuring input must be galvanically separated from the processing unit. A common realization is shown in **Figure 9**. Besides the use of SFP-modules, the galvanic separation can also be performed by capacitively or inductively couplers.

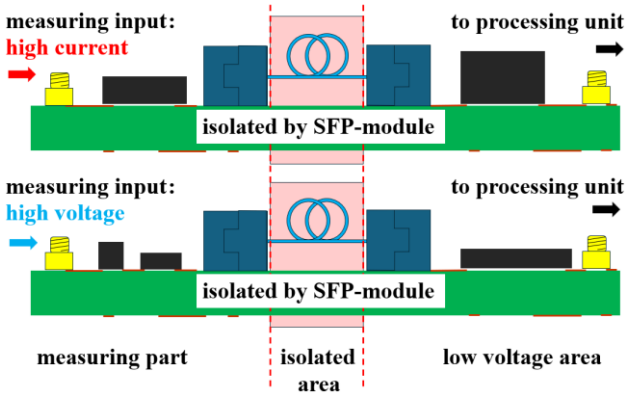


Figure 9 View of current and voltage measuring channel with isolating fiber

The requirements for power analyzers are constantly increasing. First, there is the rising voltage. The current standard for electrical measuring devices and lab equipment considers voltages up to 1.5 kV (DC) now [9]. Further challenges can be expected in continued development of certain components like WGBs (Wide Bandgap Semiconductors) or at restructuring of the energy nets towards HVDC (High Voltage Direct Current) with the aim of low loss transmission between the producer of energy and consumers [10]. In addition to the required voltage strength, another focus lies on a good synchronization of current and power measurement. For example, the runtime must not exceed a few nanoseconds between the current and the voltage measuring input [11]. In order to satisfy both the required galvanic isolation and a high synchronization, the use of a GigaCoupler holds great potential for improvement. The construction with GigaCoupler can be seen in **Figure 10**.

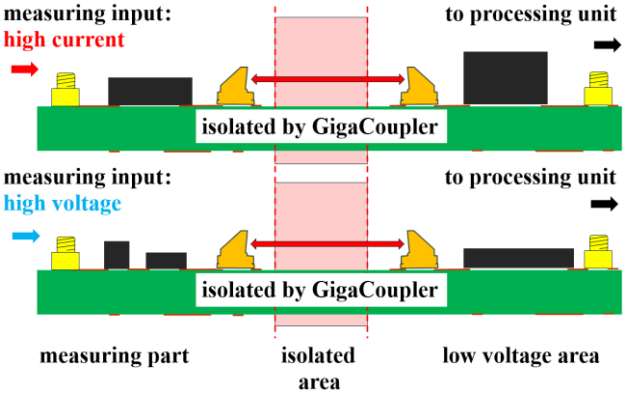


Figure 10 View of current and voltage measuring channel with GigaCoupler

A total of 4 transceivers is required for both measuring channels. The distance between these depends on the isolation strength. Since this is in the single-digit range, an air gap of 1 cm to 2 cm is sufficient.

3.3 Electro-optical circuit carrier

For three decades now, it has been announced that the optical printed circuit board would be ready to take over the entirely electrical boards as soon as the physical limits of copper have been reached. The principal structure of an electro-optical board is depicted in **Figure 11**.

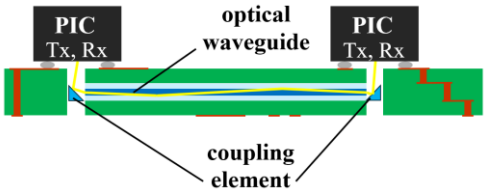


Figure 11 Cross-sectional view of a circuit carrier with electrical and optical interconnections

In addition to optical intra-board solution, a transmission of light via optical waveguides is also conceivable from board to backplane or from board to optical fiber and vice versa. This is exemplary indicated in **Figure 12**.

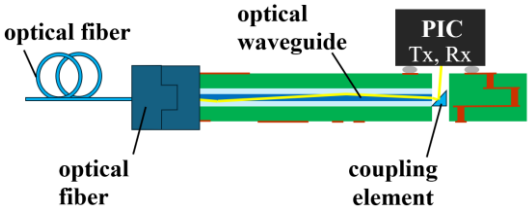


Figure 12 View of an electro-optical circuit carrier with fiber connection

The optical approach might work but comes with challenges and drawbacks. On the one hand, the limits of the copper tracks were repeatedly raised. On the other hand, the integration of an optical layer in a printed wiring board comes with disadvantages like partially production under clean room conditions, a rather exotic layout process, the necessity of using special chips with optical I/Os (PIC = Photonic Integrated Circuit), additional circuit elements for

beam deflection/ coupling, the thermal mismatch between pure polymer-layers and the conventional multilayer stack and so on.
The intended optical paths inside the multilayer structure could be replaced by the GigaCoupler, too. This approach would be able to bypass the above-mentioned drawbacks appropriately. There is a need for higher bandwidths in applications like data center racks with a second focus on low power consumption or flight computers with an optical transmission via plug-in PCBs and backplane [12]. Usually, several optical connections in parallel are required. A slot for graphic cards is typically connected with 16 lanes for example. Although these 16 interconnections could be replaced by 2x 16 transceivers, the approach might be costly. A promising alternative could be multi-channel GigaCoupler transceivers.

3.4 Slip ring replacement

Mechanical slip rings are widely used because they are inexpensive, but they are limited in data rate and face wear. The optical rotational transmission via Fiber Optical Rotary Joints (FORJ) is expensive and requires complex modules. The available capacitive type [13] doesn't require optical fibers, but poses other challenges. The small distance between rotating and non-rotating part of slip ring requires a sophisticated assembly and the sensitivity to electromagnetic interference must be viewed critically.
The GigaCoupler can replace the existing contactless solutions. Compared to FORJ, the GigaCoupler would be more cost-effective and in relation to capacitive slip rings the mentioned drawbacks can be eliminated. This gives the end user a simpler integration if only conventional (electrical) connectors are required and a conversion from electrical to optical and back is not needed.
Figure 13 shows a possible configuration for an on-axis application. The figure shows an arrangement with a rotation of 0° and 180°.

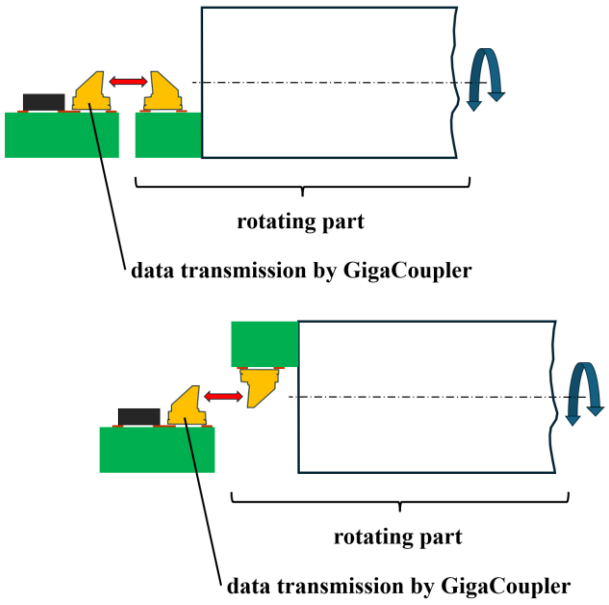


Figure 13 Example of a GigaCoupler arrangement for rotating application

Requirements for a higher bandwidth can be met by further transceiver couple(s).

4 Summary and Outlook

With the implementation of optical transceivers in parallel to a PCB shown here, it is possible to transfer data at a baud rate of 1.25 Gbps and potentially more with the next generations. Optocouplers cannot compete with this. Further advantages of the GigaCoupler lie in tracking resistance and achievable isolating voltage. Due to the separation into two transceivers, these are much higher than what can be obtained from an optocoupler. These benefits and the small form factor will be of interest for applications which have reached their limits here so far. The special lens design and the concept of GigaCoupler are protected by several IP rights.
This works shows that several applications can benefit from the GigaCoupler technology. The substitution of the fiber in optical probes results in a significant reduction in space if controller and probe are merged in a single housing. Power analyzers can benefit from isolation and high bandwidth. Excessive efforts and exotic layer structure and assembly are avoided by eliminating data transmission via embedded waveguides in PCBs. Finally, it is worth highlighting the rotational capability that enables further application fields.
Next steps include an optical characterization and a qualification of GigaCouplers behavior under rough environmental conditions. A data rate of 1 Gbps is not the upper limit. The target is 10 Gbps via an intermediate step with 5 Gbps.

5 Acknowledgment

The development of the concept for the optical communication link with beam deflection was financially supported by the Federal Ministry of Research, Technology and Space (BMFTR) under the funding mark: 16ME0366.

6 References

[1] Würth Elektronik eiSos GmbH & Co. KG, *ANO007 | Grundlagen zu Phototransistor-Optokopplern.* [Online]. Available: https://www.we-online.com/ANO007 (accessed: Nov. 7 2025).
[2] Skyworks Solutions, Inc., *Datasheet: OLI500: Miniature High CMR, High-Speed Logic Gate Optocoupler for Hybrid Assembly.* [Online]. Available: https://d1ehax0mqsd4rz.cloudfront.net/-/media/SkyWorks/Documents/Products/2701-2800/OLI500_202336H.pdf?rev=fc394978e0bd436ca4ef512224f09968 (accessed: May 20 2025).
[3] Toshiba Electronic Devices & Storage Corp., *Datasheet: TLP2367.* [Online]. Available: https://toshiba.semicon-storage.com/info/TLP2367_datasheet_en_20171025.pdf?did=36692&prodName=TLP2367 (accessed: May 17 2025).

[4] HVP High Voltage Products GmbH, *Datasheet: OPC10M.* [Online]. Available: https://www.hvproducts.de/wp-content/uploads/2021/04/d-opc10m_rev_1.0.pdf (accessed: Nov. 7 2025).
[5] FS Inc., *Specification: 10G SFP+ ACTIVE OPTICAL CABLE (AOC).* [Online]. Available: https://resource.fs.com/mall/doc/20231103163008drhhsg.pdf (accessed: Nov. 7 2025).
[6] L. L. Molex, *Datasheet: 747370009.* [Online]. Available: https://www.mouser.de/ProductDetail/Molex/74737-0009?qs=x6EjVpvqMVMWvI4k2IE8%2Fg%3D%3D (accessed: Nov. 7 2025).
[7] K. Schmieder, T. Schneider, and R. Kirrbach, "Demonstration of a Novel Gigabit Optocoupler with Small Footprint and High Isolation Voltage," in *2025 International Conference on Software, Telecommunications and Computer Networks (SoftCOM), Split, Croatia, 2025*, pp. 1–6.
[8] Rohde & Schwarz GmbH & Co. KG, *Product brochure: R&S®RT-ZISO isoliertes Tastkopfsystem.* [Online]. Available: https://scdn.rohde-schwarz.com/ur/pws/dl_downloads/pdm/cl_brochures_and_datasheets/product_brochure/3672_9996_11/RT-ZISO_bro_de_3672-9996-11_v0300.pdf (accessed: Nov. 7 2025).
[9] Gantner Instruments GmbH, *Guide: Die richtige Wahl der Messleitung für HV-Module.* [Online]. Available: https://knowledge.gantner-instruments.com/hubfs/Gantner_Instruments_Measurment_lead_German.pdf (accessed: Nov. 7 2025).
[10] energie-experten.org, *1.500-Volt-PV-Technologie auch in Deutschland wirtschaftlich.* [Online]. Available: https://www.energie-experten.org/news/1500-volt-pv-technologie-auch-in-deutschland-wirtschaftlich (accessed: Nov. 7 2025).
[11] ZES Zimmer GmbH, *Datasheet: LMG671.* [Online]. Available: https://www.zes.com/de/Produkte/Praezisions-Leistungsmessgeraete/LMG671 (accessed: Nov. 7 2025).
[12] Vario-Optics AG, *Optical PCBs: Embedded Waveguides solving Photonic Chip Assembly Bottlenecks.* [Online]. Available: https://www.swissphotonics.net/libraries.files/Board-level_photonics1.pdf (accessed: Nov. 7 2025).
[13] GAT - Gesellschaft für Antriebstechnik mbH, *Product brochure: Schleifring ROTOCAP.* [Online]. Available: https://www.gat-mbh.de/produkte/schleifringe/rotocap/ (accessed: Nov. 7 2025).

DOI 10.5162/iCCC2026/6.3

Eine dezentrale, Mesh-basierte Plattform zur Überwachung von Nutztieren für ländliche Gebiete mit eingeschränkter Konnektivität

Milan Ferus-Comelo
Duale Hochschule Baden-Württemberg
Stuttgart, Deutschland
milan.ferus-comelo@lehre.dhbw-stuttgart.de

Fabian Galka
Duale Hochschule Baden-Württemberg
Stuttgart, Deutschland
fabian.galka@lehre.dhbw-stuttgart.de

Daniel Strauß
Duale Hochschule Baden-Württemberg
Stuttgart, Deutschland
daniel.strauss@lehre.dhbw-stuttgart.de

***Zusammenfassung*—In den letzten Jahren ist die Nachfrage nach kostengünstigen, nachhaltigen und lokalen Lösungen zur Nutztierüberwachung stark angestiegen. Grund dafür sind die steigenden Kosten durch Tierverluste, die durch mangelnde Gesundheitsüberwachung, Raubtierrisse, Diebstähle oder Fluchtversuche verursacht werden. Farmlink bietet Landwirten eine moderne, digitale Überwachungslösung, die unabhängig von Mobilfunknetzen selbst in den entlegensten Regionen der Welt effizient eingesetzt werden kann. Das Weidemanagement-System implementiert ein selbstheilendes LoRa-Mesh-Netzwerk mit energieeffizientem Power-Management für einen monatelangen Betrieb.**

***Index Terms*—LoRa, Meshtastic, Herd-Monitoring, Tracking, GPS, Tierüberwachung, Landwirtschaft, IoT**

I. Einführung

FÜR landwirtschaftliche Betriebe ist eine zuverlässige Überwachung von Nutztieren entscheidend, um Tierverluste durch Krankheiten, Flucht, Raubtiere oder Diebstahl zu vermeiden und gleichzeitig die Anforderungen an Tierwohl und Wirtschaftlichkeit zu erfüllen. Studien zeigen, dass solche Verluste weltweit zu erheblichen wirtschaftlichen Schäden führen: In den USA beliefen sie sich im Jahr 2015 auf rund 3,9 Milliarden US-Dollar, was einem Anstieg von 800 Millionen US-Dollar im Vergleich zu 2005 entspricht [1]. Auch in Europa nehmen Risiken, wie steigende Raubtierbestände, wieder zu [2], [3]. Zudem verschärfen zunehmende Fälle von Viehdiebstahl die Situation: In Großbritannien wurden 2020 Schäden von etwa 2,7 Millionen Euro gemeldet [4] und im Dartmoor-Nationalpark wurden einem Landwirt innerhalb eines Jahrzehnts über 4.500 Schafe entwendet [5]. Parallel dazu bestehen weiterhin strukturelle Herausforderungen im Tierwohlmanagement, die eine regelmäßige und zuverlässige Kontrolle der Tiere erforderlich machen [6].

In der Praxis sind konventionelle Schutzmaßnahmen [7] wie Herdenschutzhunde, Zäune, regelmäßige manuelle Kontrollen oder die akustische Lokalisierung über Tierglocken [8] weit verbreitet. Sie stoßen jedoch hinsichtlich Wirksamkeit, Kosten und Arbeitsaufwand an Grenzen. Insbesondere auf weitläufigen Weiden können Tiere leicht entlaufen, unbemerkt erkranken oder gestohlen werden, während eine lückenlose manuelle Überwachung für viele Betriebe kaum realisierbar ist.

Solche Situationen führten zu einem starken Wachstum auf dem Markt für Nutztierüberwachung. Es wird vorhergesagt, dass das Marktvolumen in Europa und den USA in den nächsten acht Jahren um 8% [9] bzw. 12 % [10] steigen wird. Die Nachfrage nach digitalen und smarten Überwachungslösungen wird in den kommenden Jahren weiter zunehmen.

II. Aktuelle Ansätze der Tierüberwachung

Landwirten stehen viele unterschiedliche Methoden und Systeme zur Verfügung, die ihnen digitale und smarte Lösungen zur Nutztierüberwachung bieten. Diese Systeme werden in zwei Gruppen eingeordnet: Gesundheitsüberwachung und Tierlokalisierung [11].

Die Gesundheitsüberwachung dient bei Nutztieren der Vorbeugung von Krankheiten und Verletzungen, um im Bedarfsfall schneller reagieren zu können und diese im besten Fall zu verhindern. Bei Rindern gibt es Methoden, mit denen sich mithilfe von Bolus-Sensoren im Körper die Vitalwerte erfassen lassen [12]. Die erfassten Daten werden mit Long Range (LoRa) an Referenzknoten weitergeleitet, die in einem Mesh-Netzwerk verbunden sind. Einer dieser Knoten übermittelt die Daten dann mittels 5G. Die Herausforderung dieses Systems ist die reduzierte Signalreichweite der Sensoren, da Körperflüssigkeiten die Datenübertragung per LoRa erschweren. Die limitierten Reichweiten ermöglichen somit hauptsächlich die Überwachung in der Stallhaltung. Eine andere Methode, um die Vitalwerte der Tiere zu erfassen, ist die Erfassung der Umgebungswerte in Echtzeit mit Sensoren [13]. Übertragungsarten wie Mobilfunk und LoRa ermöglichen sehr weite Reichweiten und machen das System somit für die Weide- und Stallhaltung effektiv.

FarmLink ist ein Projekt, das aus dem Studentenwettbewerb *Competition of Students in Microsystems Applications (COSIMA) 2025* hervorgegangen ist und bei diesem bundesweiten Wettbewerb den dritten Platz erreichte. COSIMA 2025 fand im Rahmen des Microsystemtechnik (MST)-Kongresses 2025 in Duisburg statt. Die drei Autoren waren zu Beginn des Wettbewerbs an der Duale Hochschule Baden-Württemberg (DHBW) Stuttgart tätig und absolvierten dort ihren Bachelor of Engineering in Embedded Systems.

DOI 10.5162/iCCC2026/6.3

Tierlokalisierungsmethoden dienen der schnellen Ortung und helfen dabei, hohe Anschaffungskosten für Zäune oder Herdenschutzhunde zu vermeiden [11]. Meist wird Tierlokalisierung mit einem Tracker gelöst, der am Tier befestigt wird. Je nach Typ erhalten die Tiere unterschiedliche Arten von Trackern, die für verschiedene Zwecke ausgelegt sind. Halsbandtracker mit Solarfunktion [14] können über Satellitennavigationssysteme wie Global Positioning System (GPS) ihre Position erfassen und mithilfe von Bewegungssensoren einige Daten zur Tiergesundheit sammeln. Alle diese Tracker übermitteln die Daten über Mobilfunk in Kombination mit Bluetooth. Das vermeidet die Notwendigkeit von Basisstationen, allerdings macht es die Tracker von Regionen mit guter Mobilfunkabdeckung abhängig. Eine andere Möglichkeit ist die Positionsbestimmung mittels GPS als Eartag am Tier [15]. Tiere in der Umgebung, die nicht mit einem GPS-Modul ausgestattet sind, können jedoch über Bluetooth eine approximative Positionsschätzung erhalten. Diese Tracker nutzen LoRa, um ihre Daten in ein lokales LoRa-Netzwerk zu übermitteln. LoRa ermöglicht die Netzwerkkommunikation auch in weit abgelegenen Gebieten ohne Mobilfunk und macht somit die Positionsbestimmung in abgelegenen Weiden möglich. Da GPS sehr energiehungrig ist, gibt es auch weitere Lokalisierungsmethoden wie Received Signal Strength Indicator (RSSI) [16], die häufig für Indoor-Positionierung genutzt wird. Ein solches System nutzt Sensoren am Tier, welche die erfassten Daten per Funk in ein Wireless Sensor Network (WSN) übertragen [11].

Diese Beispiele sind nur einige der verschiedenen Möglichkeiten und Ansätze, um Nutztiere zu überwachen, zu lokalisieren und bei Gesundheitsänderungen rechtzeitig zu reagieren. Es gibt jedoch viele Herausforderungen bei mobilen Geräten am oder im Tier. Der Energieverbrauch ist einer der Hauptschwerpunkte, die es zu verbessern und zu optimieren gilt. Ein reduzierter Energieverbrauch garantiert eine längere Laufzeit des Geräts [11]. Dies ist vor allem in abgelegenen Regionen wichtig, in denen die Tiere, wie in den Alpen, den ganzen Sommer auf der Weide bleiben. Bei einer großen Anzahl an Tieren mit Geräten kann es sehr aufwändig werden, jedes Gerät aufzuladen oder die Batterien zu tauschen. Wireless Local Area Network (WLAN)-Netzwerke haben eine begrenzte Reichweite, daher setzen viele Produkte auf Mobilfunk [11]. Allerdings benötigen diese Systeme eine gute Mobilfunkverbindung. Andere Netzwerksysteme wie LoRa sind deutlich energieeffizienter [11] und können eigene lokale Netze aufbauen [17]. Unser Projekt *FarmLink* kombiniert mehrere dieser Ansätze und optimiert die lokale Konnektivität und Datenübertragung der einzelnen Tracker.

III. FarmLink

Das Ziel von FarmLink ist es, ein praxisnahes, bezahlbares und datenschutzfreundliches Produkt zu entwickeln. Mithilfe eines dezentralen LoRa-Mesh-Netzwerks können Daten auch ohne Mobilfunk oder zentrale Gateways übertragen werden. Dadurch ist es möglich, Informationen über den Standort und das Verhalten der Tiere auch in abgelegenen Regionen zu erfassen. Die gewonnenen Daten dienen nicht nur der Ortung, sondern ermöglichen auch neue Erkenntnisse über das Verhalten der Tiere, die Nutzung der Weiden und den Gesundheitszustand.

A. Verwendete Technologien

1) LoRa: LoRa ist ein energieeffizientes Funkverfahren auf Basis von Chirp Spread Spectrum (CSS). Es wurde für die Übertragung kleiner Datenmengen über große Entfernungen entwickelt. Es nutzt lizenzfreie Sub-GHz-Frequenzen, vor allem 868 MHz in Europa und erreicht Reichweiten von bis zu 15 km bei sehr niedrigen Datenraten. Aufgrund des geringen Energieverbrauchs können Geräte jahrelang batteriebetrieben arbeiten, was LoRa besonders für Anwendungen wie Umweltmonitoring und Landwirtschaft interessant macht [17].

2) Meshtastic: Bei Meshtastic handelt es sich um ein Open-Source-Projekt, das die LoRa-Funktechnik mit einer flexiblen Firmware nutzt, um stromsparende Mesh-Netzwerke zu realisieren. Die Software läuft auf embedded Mikrocontroller und ermöglicht es jedem Knoten, Daten zu empfangen und weiterzuleiten. Dadurch entstehen weitreichende, robuste und einfach skalierbare Netzwerke. Der Datenaustausch ist verschlüsselt, die Konfiguration erfolgt komfortabel per App und die offene Firmware erlaubt individuelle Anpassungen [18].

B. Systemarchitektur

Die FarmLink-Architektur, in Abbildung 1 dargestellt, basiert auf einem dreischichtigen Modell aus GPS-Trackern, einer Basisstation und einem Cloud-Backend. Die Tracker erfassen Positionsdaten, die von der Basisstation gebündelt und analysiert werden. Optional übernimmt die Cloud die Analysen und Visualisierung.

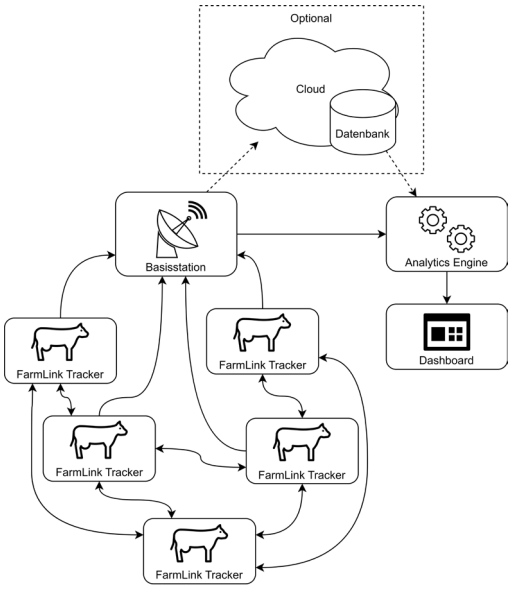


Abbildung 1. FarmLink Systemarchitektur

Das System nutzt ein dezentrales LoRa-Mesh-Netzwerk auf Basis von Meshtastic. Jeder Tracker arbeitet dabei als eigener

DOI 10.5162/iCCC2026/6.3

Knoten und Repeater. Dadurch entsteht eine selbstheilende, ausfallsichere Struktur ohne zentrale Abhängigkeiten. Geräte in der Nähe der Basisstation senden direkt, weiter entfernte Tracker kommunizieren über mehrere Hops.

Die Tracker erfassen in vordefinierten festen Intervallen GPS-Daten und übertragen diese an die Basisstation. Die Basisstation aggregiert die Pakete und speichert sie lokal ab. Die Analyse und Visualisierung der Daten kann direkt vor Ort erfolgen, sodass die Daten vollständig beim Landwirt verbleiben und die Privatsphäre gewahrt bleibt. Optional können die Daten per gesicherter 5G-Verbindung in die Cloud übertragen werden, um weiterführende Analysen wie Mustererkennung und Vorhersagen zu ermöglichen.

Mehrere Routing-Pfade gewährleisten die Datenübertragung auch bei Ausfällen von Knotenpunkten. Die Basisstation kann Daten über einen Zeitraum von mehreren Wochen lokal puffern und synchronisiert sie automatisch nach Netzunterbrechungen.

Tracker lassen sich ohne zusätzliche Konfiguration hinzufügen und integrieren sich selbstständig ins Mesh-Netzwerk. Größere Herden können mit mehreren Basisstationen abgedeckt werden. Die modulare Software ermöglicht die Erweiterung um zusätzliche Sensoren und Edge-Computing-Funktionen.

C. Tracker

Die Tracker basieren auf einem energieeffizienten nRF52-Mikrocontroller mit einem Schlafstrom von unter 3 µA [19]. Kurze Aktivphasen für GPS-Fixes und LoRa-Übertragungen ermöglichen eine sehr niedrige Durchschnittsleistung und somit einen monatelangen Betrieb. GPS und Funk werden adaptiv gesteuert, um Fix-Zeit, Reichweite und Energieverbrauch zu optimieren. Die Konstruktion der Tracker ist in Abbildung 2 dargestellt.

Abbildung 2. FarmLink Tracker

Das Gehäuse ist witterungsbeständig und robust ausgeführt. Es besteht aus zwei verschraubbaren, mittels additiver Fertigung hergestellten Halbschalen, die die Elektronik, die Antenne und den Akku sicher aufnehmen. Seitliche Schlitze ermöglichen die direkte Befestigung am Halftergurt, ohne dass zusätzliche Hardware benötigt wird. Die Konstruktion erleichtert zugleich Wartung und Batteriewechsel.

D. Basisstation

Die Basisstation bildet das zentrale Gateway zwischen dem LoRa-Mesh im Feld und der Cloud und ist somit ein zentraler Bestandteil der FarmLink-Architektur. Sie empfängt die von den Trackern über das dezentrale LoRa-Mesh-Netzwerk gesendeten Positionsdaten, bündelt und speichert diese lokal. Die Analyse und Visualisierung der Daten kann direkt vor Ort erfolgen, sodass die Daten vollständig beim Landwirt verbleiben und die Privatsphäre gewahrt bleibt. Durch die lokale Speicherung und Analyse der Daten bleibt das System auch bei temporären Netzunterbrechungen autonom und synchronisiert die Daten nach Wiederherstellung der Verbindung automatisch mit der Cloud.

Als Supernode im Mesh-Netzwerk übernimmt die Basisstation auch die Funktion eines Repeaters und trägt zur Netzstabilität und Ausfallsicherheit bei. Sie kann mehrere Routing-Pfade unterstützen und gewährleistet die Datenübertragung auch bei Ausfällen einzelner Knoten.

Optional übernimmt sie zusätzliche Funktionen wie lokale Warnungen, die Ansteuerung von Signalleuchten oder die Integration in betriebliche Managementsysteme. Aufgrund ihrer modularen Softwarearchitektur ist sie erweiterbar, z.B. um Machine Learning (ML)-basierte Vorfilterung oder direkte Kommunikation mit mobilen Geräten über WLAN oder Bluetooth. Die Kombination von LoRa-Mesh und 5G-Backhaul ergibt ein zweistufiges Kommunikationssystem, das Reichweite und Robustheit vor Ort mit einer leistungsfähigen Anbindung an die Cloud verbindet.

E. Dashboard

Abbildung 3 zeigt die Frontend-Schicht, die einen responsiven Zugriff auf alle Echtzeit- und Analysedaten ermöglicht. Für die grafische Benutzeroberfläche kommt Angular v20 als Framework zum Einsatz. Die Anwendung ist als Progressive Web App (PWA) umgesetzt und läuft direkt im Browser, wodurch sie auch offline genutzt werden kann.

Die Analysearchitektur ist modular aufgebaut. Eine lokale Engine übernimmt Echtzeitfunktionen wie Geofencing, Abweichungs- und Inaktivitätserkennung, während eine optionale Cloud-Komponente aggregierte Zeitreihen für prädiktive Analysen und Heatmaps verarbeitet.

Das Dashboard ist um eine Live-Kartenansicht zentriert, in der jedes Tier als dynamisch aktualisiertes Symbol erscheint. Mithilfe des Geofence-Tools können virtuelle Grenzen gesetzt und Warnmeldungen bei Grenzverletzungen ausgelöst werden. Historische Routen und Heatmaps lassen sich für bestimmte Zeiträume abrufen.

Zu den weiteren Funktionen gehören eine Alarmverwaltung sowie Echtzeitbenachrichtigungen bei auffälligem Verhalten. Über die Weboberfläche können Landwirte diese Meldungen ortsunabhängig abrufen.

IV. Ausblick

FarmLink befindet sich noch in der frühen Prototypenphase. Grundlegende Tests haben jedoch bereits die Funktionalität der GPS-Ortung und der Datenübertragung zwischen mehreren Geräten belegt. Die derzeitige Laufzeit eines Trackers beträgt etwas mehr als einen Monat.

Als Nächstes sollen diese Tracker in einem realen Umfeld auf einer Weide mit vielen Tieren angewendet werden, um

DOI 10.5162/iCCC2026/6.3

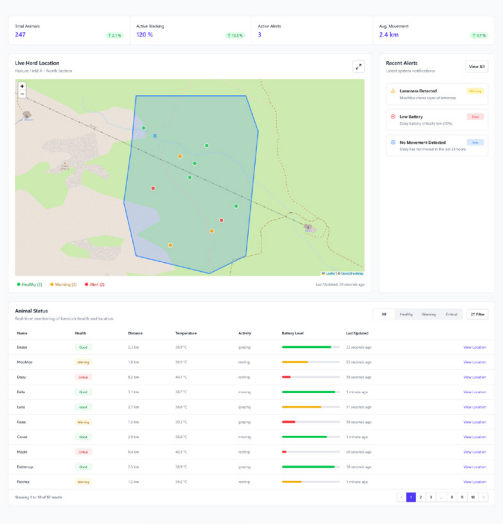


Abbildung 3. FarmLink Dashboard

mögliche Engpässe bei der Datenübertragung, der Signalreichweite und der Laufzeit festzustellen. Insbesondere muss die Belastung des Mesh-Netzwerks während der kontinuierlichen Datenübertragung von sehr vielen Geräten geprüft werden. Durch die praktische Anwendung ergeben sich auch Optimierungen der Tracking-Performance und Analysequalität. Mit den nächsten Prototypen wird bereits an Verbesserungen der Tracker geforscht. So sollen zusätzliche Sensorfunktionen für die Bewegung der Tiere integriert werden. Derzeit besitzt der Tracker nur einen Akku mit begrenzter Laufzeit. Hier kommen Energieharvesting-Methoden, wie Solar, in Frage, um manuelle Ladezyklen so minimal wie möglich zu halten.

V. Zusammenfassung

Im Rahmen des COSIMA-Wettbewerbs 2025 wurde das bewährte Konzept der GPS-basierten Tierüberwachung aufgegriffen und zu einem fortschrittlichen Weidemanagementsystem weiterentwickelt. Das FarmLink-System kombiniert energieeffiziente LoRa-Funktechnik mit einem selbstheilenden Mesh-Netzwerk. Dadurch ist die Überwachung von Nutztieren auch in abgelegenen Gebieten ohne Mobilfunkabdeckung möglich. Die Systemarchitektur ermöglicht eine kontinuierliche Positionsbestimmung bei extrem niedrigem Energieverbrauch und gewährleistet einen monatelangen Betrieb mit einer einzigen Batterieladung. Darüber hinaus wurde ein Dashboard entwickelt, das Echtzeitmonitoring mit intelligenter Verhaltensanalyse und prädiktiven Warnfunktionen verbindet.

Literatur

[1] United States Department of Agriculture, “Death Loss Trends in the U.S. Cattle Industry: 1990-2015,” 2023. besucht am 11. Nov. 2025. Adresse: https://www.aphis.usda.gov/sites/default/files/cattle-death-loss-1990-2015-info.pdf.

[2] I. Reinhardt und G. Kluth, “Large Carnivore Impacts, Case Study: Saxony, Germany. Insight into a Wolf - Livestock Conflict 25 Years after Wolf Recolonization,” EU Platform for the Coexistence of People und Large Carnivores, Berlin, 2024.

[3] Dokumentations- und Beratungsstelle des Bundes zum Thema Wolf, “Bundesweite Schadensstatistik,” 2025. besucht am 17. Nov. 2025. Adresse: https://www.dbb-wolf.de/wolfsmanagement/herdenschutz/schadensstatistik.

[4] S. Wirtschaft, “Landwirte im Vereinigten Königreich beklagen Millionenschaden,” 2021. besucht am 18. Nov. 2025. Adresse: https://www.spiegel.de/wirtschaft/grossbritannien-landwirte-beklagen-millionenschaden-durch-viehdiebstahl-a-d9ebc326-5e25-49fb-9957-3ead34ff9452.

[5] C. Cox, “I’ve lost 4500 sheep to thieves on Dartmoor,” 2025. besucht am 18. Nov. 2025. Adresse: https://www.bbc.com/news/articles/cly44r4pvd5o.

[6] A. Botzki und M. Wolfschmidt, “foodwatch Report 2023, Tierleid im Einkaufskorb. Warum alle Haltungsformen Nutztiere krank machen und wie sich das ändern lässt.,” 2023. besucht am 18. Nov. 2025. Adresse: https://www.foodwatch.org/de/tierleid-im-einkaufskorb-warum-alle-haltungsformen-nutztiere-krank-machen-und-wie-sich-das-aendern-laesst.

[7] Landesschafzuchtverband Baden-Württemberg e.V., “Herdenschutz in Baden-Württemberg; Erfahrungen und Empfehlungen für die Praxis,” 2021. besucht am 19. Nov. 2025. Adresse: https://www.herdenschutz.dvl.org/fileadmin/user_upload_herdenschutz/Fachinformationen/2020_Broschuere_Herdenschutz_BW.pdf.

[8] U. Bauer, “Kuhglocken - Lebensretter oder Tierquälerei?,” 2019. besucht am 19. Nov. 2025. Adresse: https://www.aktiontier.org/artikel/kuhglocken-lebensretter-oder-tierquaelerei.

[9] Market Data Forecase, “Europe Livestock Monitoring Market Report,” 2025. besucht am 19. Nov. 2025. Adresse: https://www.marketdataforecast.com/market-reports/europe-livestock-monitoring-market.

[10] Grand View Research, “U.S. Livestock Monitoring Market (2025-2033),” 2025. besucht am 19. Nov. 2025. Adresse: https://www.grandviewresearch.com/industry-analysis/us-livestock-monitoring-market-report.

[11] V. M. T. Aleluia, V. N. G. J. Soares, J. M. L. P. Caldeira und A. M. Rodrigues, “Livestock Monitoring: Approaches, Challenges and Opportunities,” *IJEAT*, Jg. 11, Nr. 4, S. 67–76, 2022. DOI: 10.35940/ijeat.D3458.0411422.

[12] C. Brockmann u. a., “Modular 5G-Communication and Radar based Animal Health Monitoring System for Smart Live Stock Farming,” in *Mikrosystemtechnik-Kongress 2025 (MST2025)*, Duisburg, Deutschland: VDE Verlag, 2025, S. 379–381, ISBN: 978-3-8007-6614-7.

[13] Envira. “Animal Health Monitoring on Farms,” besucht am 21. Nov. 2025. Adresse: https://envira.global/animal-health-monitoring-farms/.

[14] Nofence. “Nofence Products,” besucht am 19. Nov. 2025. Adresse: https://www.nofence.no/en-us/pricing.

[15] mOOvement. “mOOvement GPS and BLE ear tags,” besucht am 21. Nov. 2025. Adresse: https://www.moovement.com.au/.

[16] L. Nóbrega, P. Pedreiras und P. Gonçalves, “SheepIT - An Electronic Shepherd for the Vineyards,” in *International Conference on Information and Communication Technologies in Agriculture, Food and Environment (HAICTA 2017)*, Chania, Griechenland, 2017, S. 379–381. Adresse: https://ceur-ws.org/Vol-2030/HAICTA_2017_paper75.pdf.

[17] LoRa Alliance. “What is LoRaWAN?” Besucht am 20. Nov. 2025. Adresse: https://lora-alliance.org/about-lorawan/.

[18] Meshtastic LLC. “Meshtastic Introduction,” besucht am 20. Nov. 2025. Adresse: https://meshtastic.org/docs/introduction/.

[19] Nordic Semiconductor. “Nordic Semiconductor Bluetooth LE,” besucht am 21. Nov. 2025. Adresse: https://www.nordicsemi.com/Products/Wireless/Bluetooth-Low-Energy.

DOI 10.5162/iCCC2026/6.4

RFID und BLE Tracking im Industriellen Umfeld

Thomas Krause, Martin Richter, Prof. Dr.-Ing. Peggy Näser
BTU Cottbus-Senftenberg, Cottbus, Germany
thomas.krause@4.0-automation.de, naeser@b-tu.de

Abstract: Die transparente Verfolgung von Materialien und Fertigungsobjekten ist eine zentrale Voraussetzung für effiziente und digitalisierte Produktionsprozesse. Da Tracking- und Tracing-Technologien wie Radio Frequency Identification (kurz: RFID), Bluetooth Low Energy (Kurz: BLE) oder Barcodes jeweils unterschiedliche Stärken und Einschränkungen besitzen, fällt es insbesondere kleinen und mittleren Unternehmen (KMU) schwer, eine geeignete Lösung für die Verfolgung von Objekten (z.B: Material, Transportbehäter, Transportmittel, WeErkzeuge, Fertigungsaufträge und Ausrüstung) auszuwählen. Um eine praxisnahe Entscheidungsgrundlage zu schaffen, wurde im Projekt ProDat ein modularer, mobiler Track & Trace-Teststand (kurz: TT-Teststand) entwickelt. Dieser ermöglicht es, verschiedene Technologien direkt im realen Fertigungsumfeld zu erproben und deren technische sowie prozessuale Auswirkungen zu bewerten. Kern des Systems ist ein leichtgewichtiges minimalistisches Manufacturing Execution System (kurz: miniMES), welches Prozessdaten erfasst, visualisiert und Auswertungen ermöglicht. Erste Anwendungen in Unternehmen zeigen, dass der Teststand eine risikoarme, flexible und verständliche Möglichkeit bietet, Trackingtechnologien zu vergleichen und in bestehende Abläufe zu integrieren. Dadurch unterstützt er KMU wirksam bei der Auswahl geeigneter Digitalisierungslösungen.

Keywords: RFID, BLE, Bluetooth, Track and Trace, Barcode, Terminal, Prozesse, Praxis, Shopfloor

Einleitung

Die Grundlage fundierter Entscheidungen und einer funktionierenden Planung sind aktuelle Informationen über den Stand der Fertigung, des Materials und der Fertigungsaufträge. Diese zur verfügung stehenden aktuellen Informationen wird in der Fachsprache oft als Transparenz bezeichnet. Diese Transparenz in der Fertigung erfordert eine durchgängige Verfolgung relevanter Objekte wie Materialien, Betriebsmittel, ortsungebundene Anlagen, Transportmittel und Fertigungsaufträge. Die zeitnahe und zuverlässige Erfassung dieser Informationen bildet eine wesentliche Grundlage für effiziente Produktionsplanung, Prozesssteuerung und Qualitätssicherung. Da manuelle Datenerfassungen fehleranfällig sind und den Anforderungen moderner Produktionssysteme oftmals nicht genügen, gewinnen automatisierte Tracking- & Tracing-Systeme zunehmend an Bedeutung.

Für die Realisierung solcher Systeme stehen unterschiedliche Technologien zur Verfügung, darunter RFID, Barcodes, BLE, Ultra Wide Band (kurz: UWB) und kamerabasierte Verfahren. Jede dieser Technologien besitzt spezifische Vor- und Nachteile hinsichtlich Genauigkeit, Kostenstruktur, Integrationsaufwand oder Robustheit. Eine allgemeingültige, optimale Lösung existiert daher nicht; vielmehr hängt die Eignung einer Technologie stets vom jeweiligen Anwendungskontext ab. Für Unternehmen – insbesondere kleine und mittlere Unternehmen (KMU) – ist es daher oftmals schwierig, ohne praktische Erfahrungen und die finanziellen Voraussetzungen eine fundierte Technologieentscheidung zu treffen.

Vor diesem Hintergrund wurde der “Mobile Track & Trace-Teststand (kurz: TT-Teststand)” entwickelt. Er ermöglicht Unternehmen, verschiedene Identifikations- und Ortungstechnologien schnell, flexibel und unter realen Einsatzbedingungen zu erproben. Neben der Bewertung technologischer Leistungsmerkmale kann dabei auch die Integration in bestehende Arbeitsabläufe sowie die Auswirkungen auf die operativen Prozesse untersucht werden. Denn bei dem Einsatz solcher Technologien geht es nicht nur darum, ob sie technisch im geplanten Umfeld funktionieren. Ein weiterer, sehr wichtiger Faktor ist die Integrationsfühigkeit in bestehende Arbeitsprozesse. Zum Beispiel ist Ein mobiler Scanner wenig hilfreich, wenn der Werker, der ihn nutzen soll, seine Hände nicht frei hat. Der Teststand bietet damit eine risikoarme und praxisnahe Möglichkeit, fundierte Entscheidungen über den Einsatz moderner Tracking- & Tracing-Lösungen zu treffen.

Stand der Technik

Track-&-Trace-Technologien gelten als zentrale Bausteine von Industrie 4.0, da sie Transparenz im Materialfluss schaffen und datenbasierte Optimierungen ermöglichen. Die Literatur zeigt, dass verschiedene Identifikations- und Ortungstechnologien wie RFID, BLE, UWB oder kamerabasierte Systeme jeweils spezifische Stärken und Schwächen besitzen. Eine universell optimale Lösung existiert nicht; vielmehr hängt die technologische Eignung stark vom jeweiligen Prozess und der Umgebung ab [1]. RFID- und IoT-basierte Identifikationssysteme

werden in der Forschung als zuverlässige Grundlage für automatisierte und skalierbare Produktionsprozesse beschrieben. Dabei spielen Aspekte wie Datenqualität, Integration in IT-Systeme und Robustheit eine entscheidende Rolle [2]. Echtzeit-Lokalisierungssysteme (RTLS) ermöglichen die kontinuierliche Verfolgung von Werkstücken, Behältern und Betriebsmitteln. Studien zeigen messbare Verbesserungen bei Suchzeiten, Transparenz und Produktionssteuerung, weisen jedoch auch auf technische und organisatorische Integrationshürden hin [3]. RFID wird außerdem in zahlreichen Studien als Schlüsseltechnologie für durchgängige Rückverfolgbarkeit in Supply Chains und Produktionssystemen identifiziert [4]. Erfolgreiche Implementierungen erfordern jedoch angepasste Prozesse, geeignete Infrastrukturen und klare wirtschaftliche Ziele [5]. Zukünftige Entwicklungen verbinden Trackingdaten zunehmend mit digitalen Zwillingen, um Simulationen und prädiktive Analysen in Echtzeit zu ermöglichen [5]. Parallel dazu zeigt ein aktueller Überblick über geolokationsbasierte Trackingtechnologien, dass multimodale bzw. kombinierte Systeme in der Praxis häufig die größte Flexibilität bieten [6]. Insgesamt verdeutlicht der Stand der Technik, dass leistungsfähige Tracking- und Identifikationslösungen verfügbar sind, ihre Praxistauglichkeit jedoch stark kontextabhängig ist. Besonders für KMU fehlt häufig eine risikoarme Möglichkeit, verschiedene Technologien unter realen Produktionsbedingungen zu testen. Genau hier setzt ein “Mobiler Track & Trace-Teststand (kurz: TT-Teststand)” an, indem er unterschiedliche Technologien flexibel kombinierbar und praxisnah erfahrbar macht.

Systemübersicht

Im Rahmen des Projekts “ProDat” wurde gemeinsam mit Projektpartnern der Bedarf nach höherer Transparenz in Produktionsprozessen ermittelt. Dabei stellte sich heraus, das die Digitalisierung der Prozesse nur ein Teil der Lösung für mehr Transparenz und Verringerung des Dokumentationsaufwandes darstellt. Ein Kernpunkt war die Identifizierung geeigneter Tracking- & Tracingtechnologien zu identifizieren. Daher wurde anstelle eines rein konzeptionellen Ansatzes das Projektziel dahingehend weiterentwickelt, einen Teststand zu entwickeln, der direkt im realen Einsatzumfeld der Unternehmen betrieben werden kann. Das System umfasst folgende Kernkomponenten:

- **miniMES** – zentrale Datenverarbeitung und Prozessabbildung
- **Eingabemasken** - Über Eingabemasken können Informationen manuell erfasst werden
- **BLE-Tracking** – Verfolgung von Objekten mittels Bluetooth-Tags
- **RFID-Tracking** – Verfolgung von Objekten mittels RFID-Labeln und -Tags

Fig. 1: Komponenten des Teststandes in der Produktionsumgebung

- **Barcode und Terminals** – manuelle bzw. optische Erfassung zur Objektverfolgung
- **E-Label** - elektronische Label zur dynamischen Kennzeichnung ohne Batterie

Diese modulare Struktur ermöglicht es, verschiedene Technologien einzeln oder kombiniert zu erproben und so systematisch zu bewerten.

miniMES

Das miniMES ist ein leichtgewichtiges, quelloffenes Manufacturing-Execution-System, das auf einem kompakten Rechner gemeinsam mit einer Datenbank und Visualisierungskomponenten betrieben wird. Es bildet den zentralen Server des Teststandes und dient dazu, Fertigungs- und Arbeitsprozesse der Unternehmen flexibel nachzubilden. Dadurch kann nicht nur die technische Funktionalität der verfolgenden Technologien getestet werden, sondern auch deren Integration in bestehende Prozessabläufe. Dabei ist keine Anbindung an die IT-Systeme der Kunden erforderlich. Mitarbeitende können unmittelbar erleben, wie sich Abläufe und Tätigkeiten unter Einsatz moderner Trackingtechnologien verändern würden.

Das miniMES kann vollständig autark betrieben werden und erfordert keine Anbindung an IT-Systeme des Kunden. Optional ist jedoch eine Integration möglich. Arbeitsumgebungen werden im Vorfeld anhand von Stammdaten und Fertigungsaufträgen konfiguriert. Beispiele hierfür sind:

- Artikel, Artikelgruppen und Stücklisten
- Arbeitsplätze und Arbeitsplatzgruppen
- Arbeitspläne und Fertigungsaufträge

So wird vor dem Einsatz beim Kunden eine Auswahl von Stammdaten in das System eingetragen, die für den geplaten Einsatz relevant sind. Dadurch können die Mitarbeiter vor Ort am System mit ihnen vertrauten Daten arbeiten

DOI 10.5162/iCCC2026/6.4

und einen Bezug zur laufenden Produktion herstellen. Auch können so, gerade laufende also aktuelle Fertigungsaufträge erfasst und verarbeitet werden.

Durch den Einsatz von Low-Code Umgebungen in Kombination mit Visualisierungsbibliotheken können einfache Programmlogiken für den Kundeninsatz angepasst werden und auch individuelle Eingabemasken und Übersichten erstellt und angepasst werden.

Ein weiterer Baustein des eingesetzten miniMES sind Werkzeuge zum Einrichten der eingesetzten Technologien, so dass die Bediener des Teststandes nicht die individuellen Oberflächen der Lösungen erlernen müssen. Das System ist außerdem so aufgebaut, dass weitere Technologien einfach angebunden werden können.

Fig. 2: Ein BLE-Tag an einer Materialkiste befestigt

manuelle Eingabemaske

Trotz aller Bemühungen die Informationen und Aktionen wärend des Fertigungsprozesses automatisch zu ermitteln und zu dokumentieren, werden eimmer auch Aktionen über Eingbaemasken erforderlich. Unabhängig vom einzusetzenden System sind immer Verknüpfungen von Hardware (Barcodes, BLE-Tags, RFID Tags) mit logischen Objekten wie Fertigungsaufträgen, Transportkisten oder Transportaufträgen erforderlich. Dafür werden an einigen Stellen Eingabemasken benötigt. Je nachdem,

- welche Information eingegeben werden muss,
- wie diese Information eingegeben werden soll (z.B. Tastatur, Auswahlmenü, Scanner) oder
- wie die Arbeitsbedingungen am Eingabeplatz sind (müssen Handschuhe getragen werden, hat der Werker die Hände frei usw.)

müssen diese Masken sehr unterschiedlich aufgebaut werden. Daher sind sie in diesem miniMES einfach individuell zu erstellen und anpassbar.

BLE Tracking

Bluetooth-Tags ermöglichen die drahtlose Verfolgung von Objekten in der Fertigung. Industrietaugliche Tags sind in verschiedenen Bauformen erhältlich und können auch in staubigen oder feuchten Umgebungen sowie im Außenbereich eingesetzt werden. Die Tags senden regelmäßig Signale aus, die von stationären Ankern empfangen werden. Durch Auswertung der Signalstärke kann das System die Position des Objekts bestimmen. Ähnlich wie bei RFID-Systemen positioniert man dazu die Anker an relevanten Positionen im Fertigungsprozess. Das können Arbeitsplätze, Durchfahrten, Lagerplätze oder Positionen am Transportsystem sein.

Zu Beginn eines Prozesses wird ein Tag einem Objekt zugeordnet und dies im miniMES hinterlegt. Nach Abschluss des Prozesses wird der Tag vom Objekt entfernt. Gleichzeitig wird die Verknüpfung des Tags zum Objekt im miniMES gelöst. Im Falle von wiederverwendeteten Transportsystemen, wie Kisten, Boxen, Paletten und weitere, verbleibt das Tag meist am Objekt. Dann wird im miniMES nur die Verknüpfung zum Objekt (Fertigungsauftrag, Materialcharche etc.) aufgehoben.

Zwischen diesen beiden Schritten erfolgt die Positionsbestimmung automatisch. Je nach Konfiguration können daraus Arbeitsgangbuchungen, Lagerbewegungen oder weitere Prozessschritte abgeleitet werden.

RFID Tracking

RFID-Label sind in der Regel passive Transponder, die von aktiven Antennen ausgelesen werden. Sie sind in vielen Varianten verfügbar, darunter

- Etiketten mit bedruckbarer Oberfläche
- kleine dünne Folien, die auf oder unter dünne Oberflächen befestigt werden können.
- hitzebeständige und/oder chemikalienresistente Label
- Label, die für die Montage auf Metalloberflächen geeignet sind.

An den relevanten Orten werden Reader mit Antennen installiert, die die RFID-Label automatisch auslesen und einer Position im Fertigungsporzess zuordnen. Aufgrund ihrer geringen Kosten eignen sich viele RFID-Labels auch für den Einmalgebrauch. Bei wiederverwendbaren Labels wird analog zu BLE-Tags ein Zuordnungs- und ein optionaler Trennprozess benötigt.

Barcode und Terminals

Barcodes und QR-Codes stellen eine weit verbreitete und kostengünstige Methode zur Objektverfolgung dar. An den relevanten Orten werden Scanner installiert oder mobile Endgeräte verwendet. Die Kennzeichnung erfolgt über Etiketten, die direkt im Teststand gedruckt werden

Fig. 3: Ofeneinsatz mit innen befestigtem BLE-Tag

können. Jede Erfassung wird mit Ortsangabe an das miniMES übermittelt. Der technologische Aufwand ist gering, jedoch entsteht ein höherer manueller Anteil im Prozess. Zudem sind optische Systeme anfälliger gegenüber Verschmutzung und Umwelteinflüssen.

Erfahrungen

Der Teststand wurde bereits erfolgreich in mehreren Unternehmen eingesetzt. Bei der Ortrander Eisenhütte GmbH konnte beispielsweise eine geeignete Lösung zur automatischen Verfolgung von Ofeneinsätzen entwickelt werden—trotz der Herausforderung, dass Tags ausschließlich im Inneren der Ofeneinsätze platziert werden konnten, da Außenflächen während der Lackierung frei bleiben müssen.

Besonders hilfreich erwiesen sich die Visualisierungen im miniMES, da sie die erfassten Positionen und Prozesszustände transparent darstellten. Mitarbeitende konnten so unmittelbar nachvollziehen, wie Trackingtechnologien funktionieren und welchen Einfluss sie auf bestehende Arbeitsabläufe haben.

Der vorgestellte Track-&-Trace-Teststand zeigt, dass verschiedene Trackingtechnologien unter realen Produktionsbedingungen erprobt werden können, ohne den laufenden Betrieb wesentlich zu beeinträchtigen. Durch seinen autarken Aufbau ist keine Einbindung in die IT-Infrastruktur der Unternehmen erforderlich. Die Testumgebung ermöglicht es, technologischen Nutzen, Prozessveränderungen und organisatorische Auswirkungen praxisnah zu bewerten.

Somit dient der Teststand als anbieterneutrales und technologieoffenes Werkzeug, um Unternehmen bei der Auswahl und Einführung geeigneter Trackinglösungen zu unterstützen und gleichzeitig Prozesse direkt vor Ort weiterzuentwickeln.

Danksagung

Dieses Projekt wurde durch das Bundesministerium für Bildung und Forschung unter Förderkennzeichen 03WIR0110A gefördert.

References

[1] F. Ahmed, M. Phillips, S. Phillips, and K.-Y. Kim, “Comparative study of seamless asset location and tracking technologies,” vol. 51, pp. 1138–1145.

[2] A. Frankó, G. Vida, and P. Varga, “Reliable identification schemes for asset and production tracking in industry 4.0,” vol. 20, no. 13, p. 3709.

[3] A. Rácz-Szabó, T. Ruppert, L. Bántay, A. Löcklin, L. Jakab, and J. Abonyi, “Real-time locating system in production management,” vol. 20, no. 23, p. 6766.

[4] B. Unhelkar, S. Joshi, M. Sharma, S. Prakash, A. K. Mani, and M. Prasad, “Enhancing supply chain performance using RFID technology and decision support systems in the industry 4.0–a systematic literature review,” vol. 2, no. 2, p. 100084.

[5] A. Ullah, M. Younas, and M. S. Saharudin, “Digital twin framework using real-time asset tracking for smart flexible manufacturing system,” vol. 13, no. 1, p. 37.

[6] R. Pares, I. Corthier, and A. L. Her, “State of the art in geolocation-based asset tracking technologies for logistics and production containers in industrial environments,” Publisher: Unpublished.

DOI 10.5162/iCCC2026/7.2

1D phased array for imaging applications in air utilizing MEMS ultrasonic transmitters

Marcel Jongmanns, Fraunhofer Institute for Photonic Microsystems (IPMS), Dresden, Germany,
marcel.jongmanns@ipms.fraunhofer.de
Jorge Mario Monsalve, Fraunhofer Institute for Photonic Microsystems (IPMS), Dresden, Germany

Abstract

We designed a 1D ultrasonic array for phased array applications based on the lateral micromachined ultrasonic transducers (L-CMUT), which are in-house designed and manufactured at the Fraunhofer IPMS. The 1D array is a pretest for a 2D configuration, which would allow beam steering in 2 dimensions. The functionality of the 1D array and its control electronics have been tested and verified. Steering of the beam is possible in 0.33° steps in a range of ±30° before the amplitude decreases up to a maximum angle of ±60°. To characterize the sound field, a setup utilizing an industrial cobot equipped with a reference microphone has been realized. To test the imaging capabilities different objects have been placed up to 1 m in front of the phased array. The measurement system can be used to detect the distance and position of an object relative to the position of the phased array.

1 Introduction

Phased arrays are commonly used for medical imaging, i.e. for measurements in water or soft tissue. Air-based ultrasonic imaging can be of interest to perform a presence detection or object tracking inside a room. It could also be used to increase the information acquired by parking sensors in cars, as it not only gives information about the distance, but also the incident angle; nonetheless, airborne ultrasound is often only used for distance measurements. One restriction that has usually limited the development of phased arrays in air is a transducer size below λ/2, where λ is the wavelength. A center-to-center distance below λ/2 is necessary to avoid so-called 'grating lobes': peaks in the radiation pattern that are comparable to the main lobe but lie at an undesired angle [1]. Bulk piezoelectric transducers usually have a membrane diameter above λ/2, although some methods exist to implement a phased array with them, such as external funnel structures [3] or spiral arrays [4] — both of which increase the size of the device significantly. MEMS components, such as capacitive MUTs (CMUTs) have been implemented for imaging purposes in air, although slightly trespassing the λ/2 limit [5]. A recent work has been published, where a MEMS microphone (which is naturally compact) is operated as a transmitter in a phased array [6], detecting objects in the centimeter range. Both the works in [5] and [6] achieve a range below 70 cm. With a lateral CMUT (L-CMUT), the concept of which we describe in [2], there is more flexibility for reducing the chip size below λ/2 without affecting neither its resonance frequency nor its bandwidth, only reducing the signal amplitude. In this work, we implement a one-dimensional array of L-CMUTs operating as transmitters and evaluate its performance.

2 Setup

A 1D array of 8 L-CMUT is mounted on a PCB as ultrasonic transmitter and 3 wideband MEMS microphones (Knowles SPH18C3LM4H-1) are used as receivers (Fig. 1). Each of the L-CMUT elements is 4 mm x 4 mm in size.

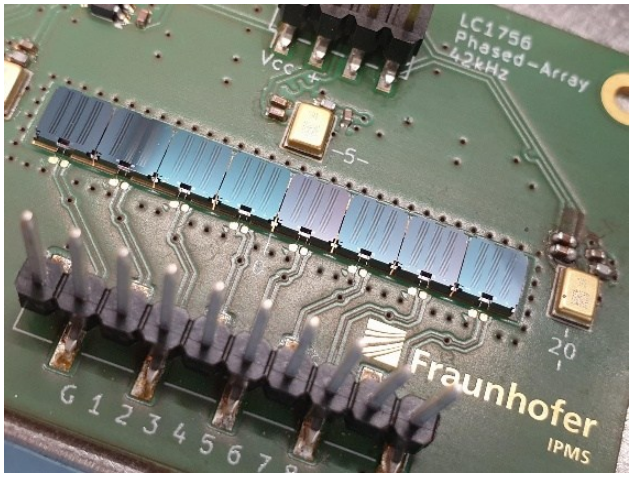


Figure 1: 1D-Array consisting of 8 L-CMUT elements and microphones as receiver.

The spacing between the center of the elements is 4.08 mm. This corresponds to a transmitting frequency of 42 kHz for phased array applications. Out of the 3 receivers only one is used in the later tests. It is in the middle of the array 5 mm (center of L-CMUT to center of acoustic port) above it. Knowing the spacing between consecutive elements (s) and the speed of sound (c), the time delay (Δt_m) that ought to be introduced the m*th* transducer (m=0, 1, …, N-1) in order to shift the main lobe by an angle ϕ can be calculated as follows [1]:

$$\Delta t_m = \begin{cases} m\frac{s}{c}\sin\phi\,, & \phi \geq 0 \\ (N-m-1)\frac{s}{c}|\sin\phi\,|, & \phi < 0 \end{cases}$$

The control electronics consists of a boost converter, output drivers, input amplifiers and a STM32 Nucleo-G474RE development board to control the system. The system is supplied with 5V via USB, which is also used to communicate with a PC. The boost converter generates the required 30 V bias voltage for the L-CMUT. The phased array is controlled by 8 timers of the microcontroller. The 3.3 V output is amplified to 30 V as AC signal for the L-CMUT, resulting in a square wave from 15 V to 45 V to

DOI 10.5162/iCCC2026/7.2

drive the L-CMUT. The microphone output is amplified by a variable factor and filtered by a hardware low-pass filter as anti-aliasing filter before the analog data is sampled at 2 MHz by the microcontroller.

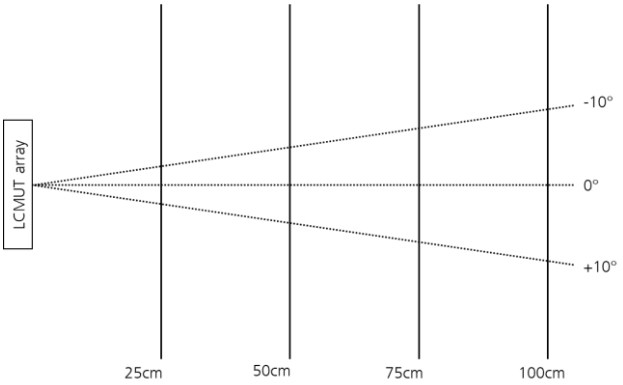


Figure 2: Schema of the field-of-view and object placement.

The laboratory test for object detection was performed on a table in a quiet office to reduce background noise. The 1D array was placed on one side of the table. Three different objects with different shapes were placed on the table. The objects were placed at 25 cm intervals from 25 cm to 125 cm apart from the measurement system (Fig. 2). For each object different angles relative to the system from -10° to +10° were tested. With this data the distance as well as the angle between the system and the object should be determined.

3 Characterization

To characterize the functionality of the phased array, a measurement setup was designed using a Universal Robots UR10 (Fig. 4). The sound pressure was measured using a calibrated GRAS 46DD 1/8" microphone, which was attached to the tool flange of the robotic arm. The UR10 was controlled by a PC running a Python script. This allowed the microphone to move in a hemisphere around the device under test while pointing the microphone at the device (Fig. 3). While this can be used to measure a 3D acoustic field, only a 2D measurement is shown. The measurement is performed along the center line along the length of the array in which the acoustic wave is steered. The angle Θ describes the angle between the array and the perpendicular line, i.e. at $\Theta = 0°$ the acoustic beam travels straight on without steering. Transmitting, receiving and moving the robot was synchronized using the Python script.

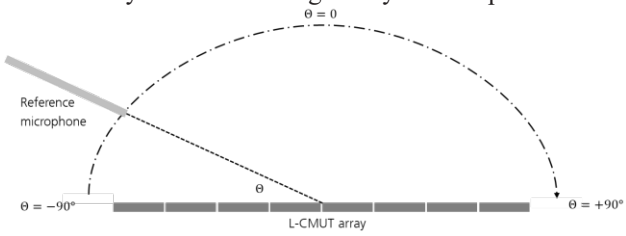


Figure 3: Schema of the movement of the reference microphone by the robot.

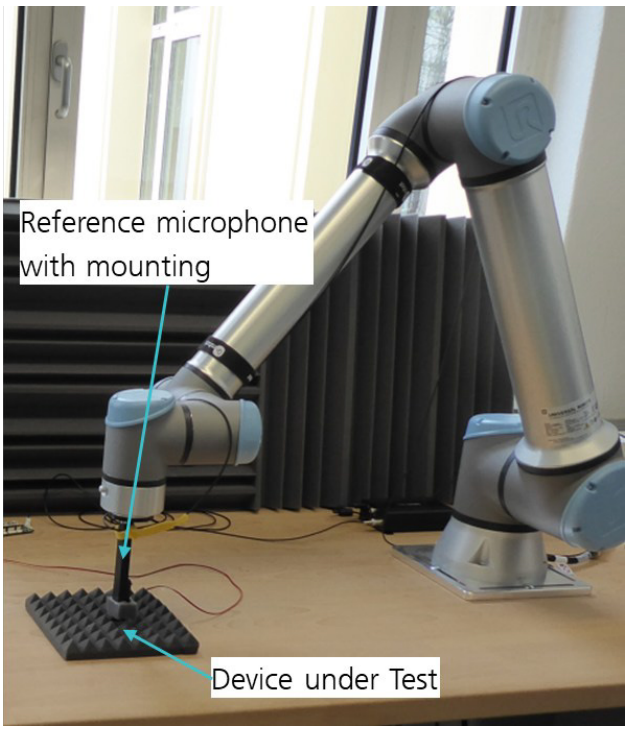


Figure 4: Actual measurement setup on the cobot.

4 Measurements

First, the array was characterized by the robotic measurement setup. The beam was steered from 0° to 90° in 5° steps. For each of these measurements the sound pressure was measured along the center line of the array in 1° steps in 10 cm distance (Fig. 5). From a steering angle of 0° to 25°, the amplitude is above 2 Pa. At 30° the amplitude falls below 2 Pa and diminishes further above 50°. Until 60° the set steering angle can be reached. At higher angles the side lobes become more prominent while the amplitude of the main beam is reduced even further, and the set steering angle cannot be reached anymore.

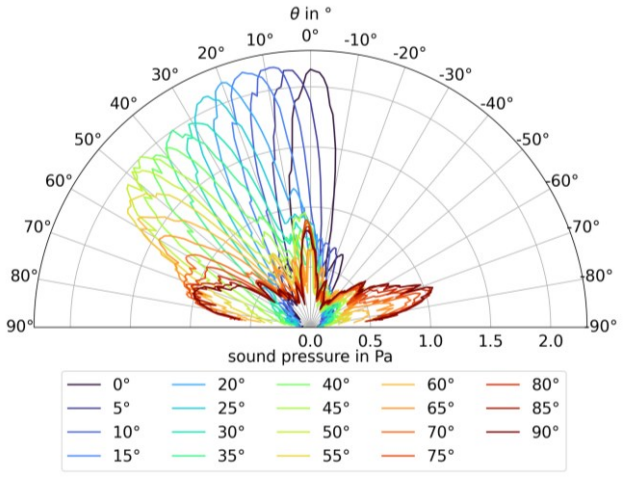


Figure 5: Measured sound pressures for different deflection angles.

DOI 10.5162/iCCC2026/7.2

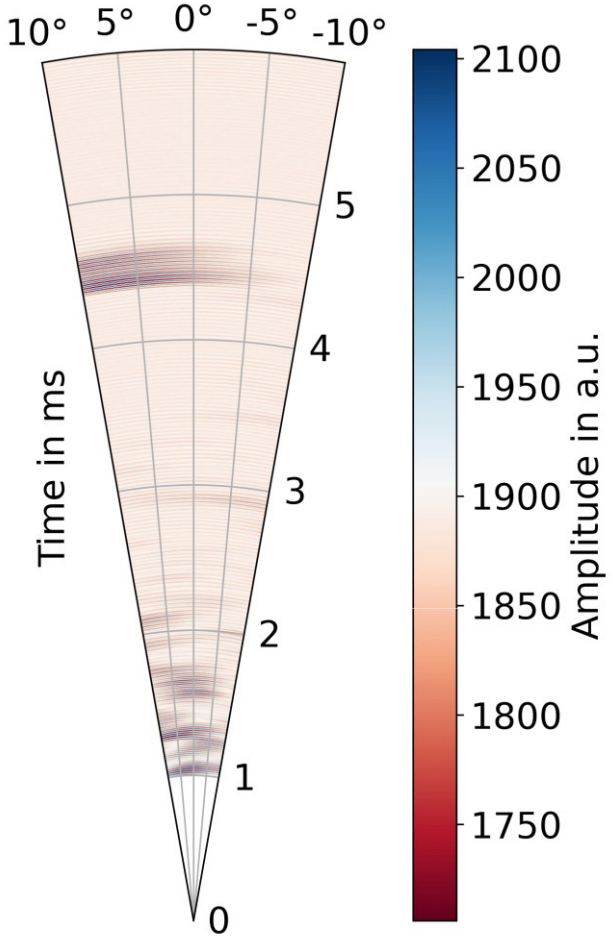


Figure 6: Line scan of an object placed 75 cm in front of the L-CMUT array at an angle of 5°.

As an example for object detection using a 1D line scan, a measurement of an object at 75 cm distance and an angle of 5° is shown in Fig. 6. The scan is performed with a steering angle from -10° to +10° in 0.33° steps. The first millisecond of the measurement has been omitted since it is governed by crosstalk. Even after that there is still a lot of noise caused by the table the system is on, the wall of the room and other objects in vicinity. However, there is also a peak in signal amplitude around 4.5 ms time-of-flight. At 75 cm distance, the total distance for a pulse-echo measurement is 150 cm. At room temperature the expected time-of-flight is $tof = \frac{1.5m}{343\frac{m}{s}} = 0.00437ms$, with the speed of sound 343m/s. This correlates well with the observed value.

Overall three different objects have been used for the scan. One was a box with a rectangular shape, one container with rectangular shape but rounded corners and one circular bottle. Looking closely at the reflected echo, one can see differences in the signal (Fig. 7). The box with sharp corners also shows a sharp difference between high and low amplitude in the reflection. The rounded corners and circular shape introduce phase shifts in the signal as well as a change in the amplitude of the reflected signal. This suggests that it is possible to not only detect the distance and position of an object, but also the shape.

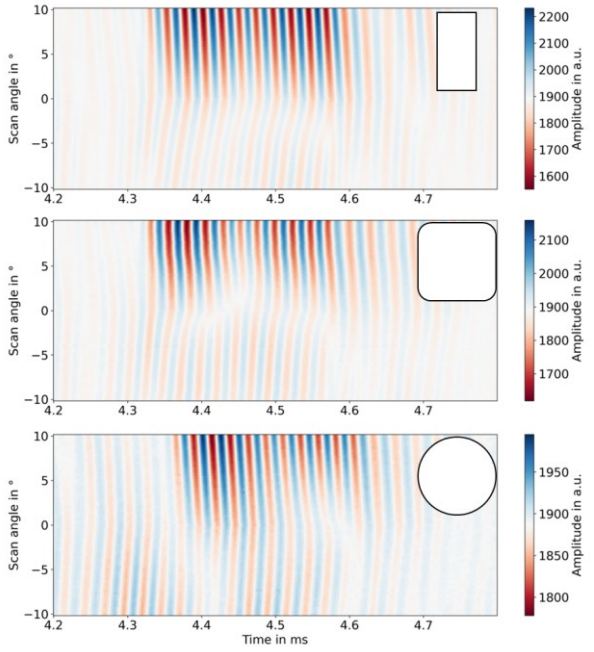


Figure 7: Reflection pattern of the scan of different object placed at 10° and 75 cm distance. Top: rectangular with sharp corners; Middle: rectangular with rounded corners; Bottom: circular.

One disadvantage of the line scan is the long measurement time. At 9 ms maximum time-of-flight, the maximum distance is around 3 m, or 1.5 m distance between measurement system and object. With a fine scan from -10° to +10° in 0.33° steps it requires 61 consecutive measurements for one image, which is overall 0.55s, excluding overhead to transfer the data from the system to the PC for further processing. A way to get around this restriction is to utilize different measurement approaches. Another tested approach tries to encode angular information by sending signals with different frequencies. As the L-CMUT is a wideband transmitter, the same setup can be used to make a pulse-echo measurement where transmitter 1 sends a burst of 15 pulses at 40kHz, transmitter 5 at 50kHz and transmitter 8 at 60kHz. Transmitters 1 and 8 are the outermost and 5 is one of the middle transmitters. The distance between the transmitters is known from the design und the distance to an object can be calculated from the time-of-flight. This information can be used to calculate the distance and angle using the law of sines. Fig. 8 shows the measured signal, which has been treated with 3 different digital bandpass filters to retrieve each of the 3 different transmitter signals.

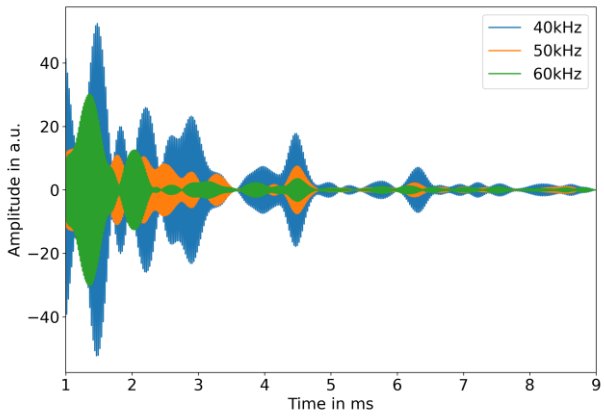


Figure 8: Decomposed signals of one multi-frequency measurement.

5 Discussion and outlook

A 1D phased array for ultrasonic line scans has been realized using L-CMUT wideband ultrasound transmitters. A custom driver electronic is used to control the system. The function of the system has been validated by making measurements of the acoustic field by an automated, robotic measurement setup.

Preliminary measurements utilizing a fine-pitched line scan with 0.33° resolution have been performed. Different objects have been placed at distances up to 1 m in front of the array and at different angles, which can be seen in the measured data. However, the system still suffers from crosstalk, which introduces a minimum distance to the object of about 50 cm. Ranges up to 1.25 m could be measured with this system during the laboratory tests, surpassing previous developments of in-air ultrasonic phased arrays. A new revision of the L-CMUT is currently being manufactured, which can be driven at higher voltages. This should increase the distance above 3m in a pulse-echo configuration.

A more extensive evaluation of possible algorithms for data processing is still open. While established algorithms like peak detection or correlation are easy to implement, they are susceptible to the crosstalk or static objects near the transmitter. One possibility is that the signal is normalized by removing static noise and only detect changes in the system. This would allow the system to detect moving objects for e.g. presence detection. Another planned approach is to utilize and evaluate AI-based algorithms, which could be used to detect different patterns in the ultrasonic signal and correlate them to different shapes of the object.

6 Acknowledgements

This work is co-funded by the German Federal Ministry of Research, Technology and Space in the project iCampus 2 (No. 16ME0420K). The authors like to thank Andreas Mrosk and Sven Püschel for support with the acoustical measurements and wire bonding of the MEMS devices.

7 Literature

[1] L. W. Schmerr Jr, Fundamentals of Ultrasonic Phased Arrays, Springer, 2015

[2] J. M. Monsalve, A. Melnikov, M. Stolz, A. Mrosk, M. Jongmanns, F. Wall, S. Langa, I. Marica-Bercu, T. Brändel, M. Kircher, H. A. G. Schenk, B. Kaiser, H. Schenk, “Proof of concept of an air-coupled electrostatic ultrasonic transducer based on lateral motion”, Sensors and Actuators A: Physical, vol. 345, no. 113813, 2022, doi: 10.1016/j.sna.2022.113813

[3] G. Allevato, J. Hinrichs, D. Grosskurth, M. Rutsch, J. Adler, A. Jäger, M. Pesavento, M. Kupnik, “3D imaging method for an air-coupled 40 kHz ultrasound phased-array”, in Proceedings of the 23rd International Congress on Acoustics, 2019, pp. 4797-4804

[4] G. Allevato et al., "Air-Coupled Ultrasonic Spiral Phased Array for High-Precision Beamforming and Imaging," in IEEE Open Journal of Ultrasonics, Ferroelectrics, and Frequency Control, vol. 2, pp. 40-54, 2022, doi: 10.1109/OJUFFC.2022.3142710

[5] K. K. Park, B. T. Khuri-Yakub, “3-D airborne ultrasound synthetic aperture imaging based on capacitive micromachined ultrasonic transducers”, Ultrasonics, vol. 53, pp. 1355-1362, 2013, doi: 10.1016/j.ultras.2013.04.003

[6] S. Anzinger et al., "Low Power Capacitive Ultrasonic Transceiver Array for Airborne Object Detection," 2020 IEEE 33rd International Conference on Micro Electro Mechanical Systems (MEMS), Vancouver, BC, Canada, 2020, pp. 853-856, doi: 10.1109/MEMS46641.2020.9056182

DOI 10.5162/iCCC2026/7.3

Ultrasound-Based Interferometric Hip Implant Monitoring: Analysis of Frequency Selection and Minimum Detectable Gap Width

Jan Lützelberger and Klaus Stefan Drese
ISAT – Institute of Sensor and Actuator Technology, Coburg University of Applied Sciences and Arts, Coburg, Germany
jan.luetzelberger@hs-coburg.de

Abstract

Early detection of aseptic or septic loosening of hip endoprostheses is clinically crucial but still relies primarily on projection radiography, which lacks sensitivity for sub-millimetre interlayers at the bone–implant interface. Recent work by Lützelberger et al. demonstrated that ultrasound interferometry enables quantification of the width of this interface by analysing frequency-domain interference patterns of reflections from the soft-tissue gap. However, several open theoretical questions remained, particularly regarding the fundamental detection limits for small gap widths and the optimal choice of excitation frequency.
This work provides a mathematical analysis of (i) the minimal resolvable gap width based on the Gabor limit, bandwidth constraints, and the signal-to-noise ratio (SNR) and (ii) the optimal operating frequency considering attenuation in soft tissue and the limited thickness of the cortical bone layer. The results show that an excitation frequency around 3 MHz represents an optimum here. Furthermore, closed-form expressions for the minimum detectable width are derived, predicting detection thresholds in the range of 156–212 µm for different SNR conditions – consistent with previously reported experimental results. These findings provide theoretical justification for system design choices in interferometric ultrasound for implant-integration monitoring.

Kurzfassung

Die Früherkennung einer aseptischen oder septischen Lockerung von Hüftendoprothesen ist klinisch von entscheidender Bedeutung, stützt sich jedoch nach wie vor in erster Linie auf Projektionsradiographie, die für submillimetergroße Zwischenräume an der Schnittstelle zwischen Knochen und Implantat nicht empfindlich genug ist. Jüngste Arbeiten von Lützelberger et al. haben gezeigt, dass die Ultraschallinterferometrie eine Quantifizierung der Breite dieser Schnittstelle durch Analyse der Interferenzmuster im Frequenzbereich der Reflexionen aus dem Weichteilspalt ermöglicht. Es blieben jedoch einige theoretische Fragen offen, insbesondere hinsichtlich der grundlegenden Nachweisgrenzen für kleine Spaltbreiten und der optimalen Wahl der Anregungsfrequenz.
Diese Arbeit liefert eine mathematische Analyse (i) der minimalen auflösbaren Spaltbreite auf Grundlage der Gabor-Unschärfe, der Bandbreitenbeschränkung des Ultraschallpulses und des Signal-Rausch-Verhältnisses (SNR) sowie (ii) der optimalen Sendefrequenz unter Berücksichtigung der Dämpfung im Weichgewebe und beschränkten Dicke des kortikalen Knochens. Die Ergebnisse zeigen, dass eine Anregungsfrequenz um 3 MHz hier einen optimalen Kompromiss darstellt. Darüber hinaus werden geschlossene Ausdrücke für die minimal erkennbare Spaltbreite abgeleitet, die Erkennungsschwellen im Bereich von 156–212 µm für unterschiedliche SNR-Bedingungen vorhersagen – in Übereinstimmung mit zuvor vorgestellten experimentellen Ergebnissen. Diese Ergebnisse liefern eine theoretische Begründung für Systemdesignentscheidungen bei interferometrischem Ultraschall zur Überwachung der Implantat-Integration.

1 Introduction

1.1 Clinical Motivation

Hip joint prostheses are among the most frequently implanted orthopedic devices, yet aseptic loosening continues to be the most common long-term complication [1]. Loosening originates from the formation of a soft-tissue layer between bone and implant – typically in the range of a few hundred micrometers to several millimeters [2]. Detecting early-stage changes in the gap width is clinically important, both for detecting loosening itself and for identifying tissue changes caused by bacterial biofilms [3]. Current diagnostics rely largely on plain radiography, which lacks the spatial resolution needed to detect gaps below 1 mm and cannot characterize the interfacial tissue [4]. Ultrasound offers a radiation-free alternative, but classical time-of-flight ultrasound imaging fundamentally struggles with thin interlayers: when the gap width is about or smaller than the acoustic wavelength, reflections from the front and back surfaces overlap in time and the interlayer thickness as a correlate to loosening cannot be determined directly in time domain [5,6,7].

1.2 Interferometric Ultrasound Approach

Lützelberger et al. [8] proposed a frequency-domain approach using the interferometric pattern generated by multiple reflections within the bone–gap–implant three-layer system. Using different idealized setups as well as a more realistic bone-implant system, the method could successfully determine water gap widths in the range of approx. 160 µm to 2 mm. While experimentally successful, the

DOI 10.5162/iCCC2026/7.3

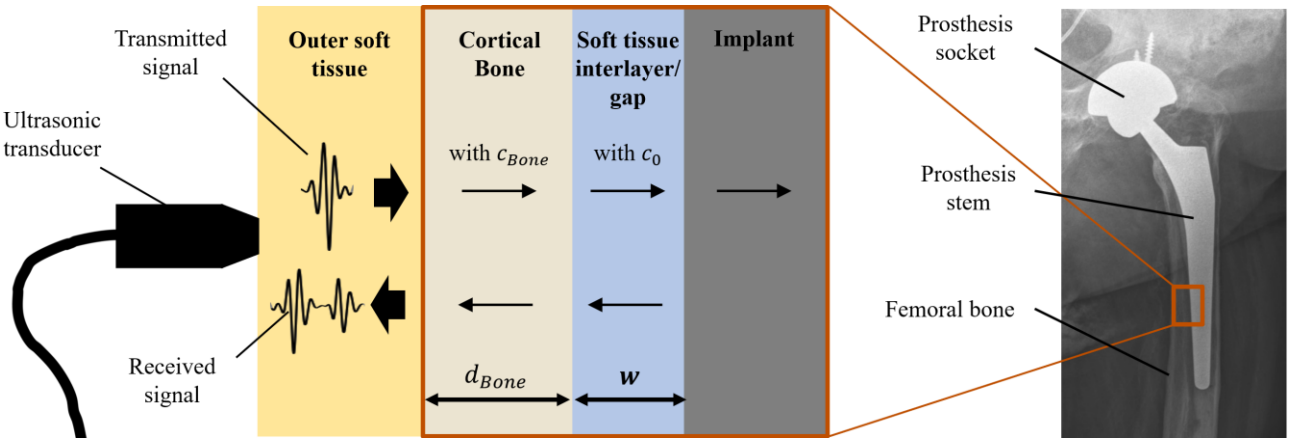


Figure 1 Multilayer model of the thigh with hip prothesis stem.

method still lacks a rigorous theoretical justification for two critical design choices:

1. What fundamental limit exists for the minimum detectable gap width, and how this limit depends on pulse length and SNR.
2. Why an excitation frequency of approximately 3 MHz is reasonable, given large anatomical variability and strong frequency-dependent attenuation.

The aim of this paper is therefore to provide a compact analytical treatment of these questions, suitable for guiding future clinical device optimization.

2 Methods

2.1 Simplified Physical Model of Ultrasonic Reflection

As Figure 1 shows, the anatomical situation of a thigh bone with a hip prosthesis stem is modelled as a layered system consisting of soft tissue (skin, fat, muscles), cortical bone, a thin soft-tissue interlayer of width w, and the implant [8,9]. For longitudinal plane waves, multiple reflections at the layer interfaces interfere, forming a frequency-dependent reflection coefficient $R(f)$ [8,10].

For gaps small compared to the wavelength, time-domain separation of the frontside and backside reflections is impossible. Instead, destructive interference produces periodic minima in the amplitude spectrum of the reflected signal. Their spectral position f_s follows from [11]

$$f_s = s\frac{c_0}{2w} \tag{1}$$

with $s \in \mathbb{Z}$ where c_0 is the sound velocity in the interlayer. The spacing of the minima therefore results from [12]

$$\delta f = \frac{c_0}{2w}. \tag{2}$$

Thus, the measurement task for determining the interlayer width w reduces to detecting the position f_s of minima within the amplitude spectrum or their spacing δf.

2.2 Bandwidth Constraints from the Gabor Limit

For time-frequency analyses in signal processing, uncertainty can be expressed by the Gabor limit [13] stating

$$\sigma_t \sigma_f \geq \frac{1}{2\pi} \tag{3}$$

for the relation between the standard deviations σ of time t and frequency f, referring to amplitude signals. For a Gaussian-shaped pulse, using the Full Width at Half Maximum (FWHM) Δ with

$$\Delta = 2\sqrt{2\ln(2)}\ \sigma \tag{4}$$

as the characteristic pulse width is more suitable because it is most straight forwardly determined from experimental data. This yields to

$$\Delta_t \Delta_f \geq \frac{4\ln(2)}{\pi} \approx 0.88. \tag{5}$$

Consequently, shorter pulses provide larger bandwidth, which increases the number of observable minima.

2.3 SNR-Dependent Detectable Gap Width

Assuming a Gaussian shaped amplitude spectrum $A(f)$ of the excitation signal (see Figure 2) with

$$A(f) = A_0\, \mathrm{e}^{-\frac{1}{2}\left(\frac{f-f_0}{\sigma_f}\right)^2} \tag{6}$$

with central frequency f_0 and maximal amplitude A_0 is a reasonable approximation for the Hanning windowed sine burst used by Lützelberger et al. [8].

To detect at least one spectral minimum in the presence of white noise with amplitude level N_0, the minimum position f_s needs to be within the interval

$$f_{N_0,-} < f_s = \frac{c_0}{2w} < f_{N_0,+}\,, \tag{7}$$

using (1), whereby $f_{N_0,-}$ and $f_{N_0,+}$ are the limits of the usable bandwidth of the transducer, i.e., the frequencies where the noise level N_0 crosses the spectrum $A(f)$ of the excitation signal.

Rearranging (7) with respect to w leads to the limits of detection of the interlayer width with

$$w < \frac{sc_0}{2f_{N_0,-}} \tag{8}$$

as the upper limit of the detectable gap width w as well as

$$w > \frac{sc_0}{2f_{N_0,+}} \tag{9}$$

DOI 10.5162/iCCC2026/7.3

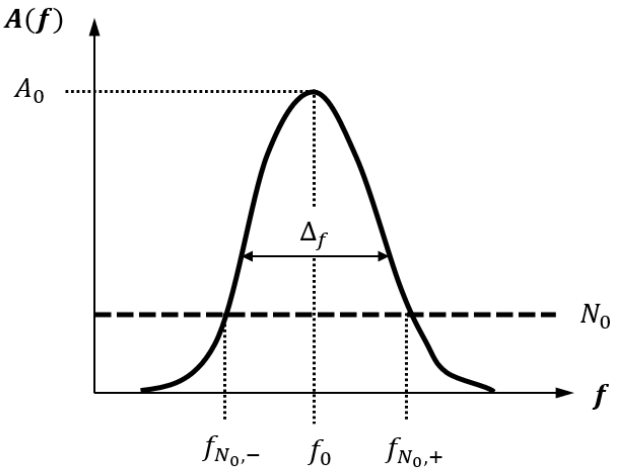


Figure 2 Schematic drawing of the amplitude spectrum $A(f)$ of the excitation signal in [8] including a white noise level N_0.

as the lower limit of the detectable gap width w.
$f_{N_0,-}$ and $f_{N_0,+}$ can be obtained using (6) and (4) leading to

$$A\left(f_{N_0}\right) = A_0\, \mathrm{e}^{-\frac{4\left(f_{N_0}-f_0\right)^2 \ln(2)}{\Delta_f^2}} = N_0. \tag{10}$$

Introducing the well-known signal-to-noise ratio SNR with

$$SNR = \frac{A_0}{N_0} \tag{11}$$

and solving (10) for f_{N_0}, using (5), gives

$$f_{N_0} = f_0 \pm \frac{2\sqrt{\ln(SNR)\ln(2)}}{\pi \Delta_t} \tag{12}$$

for $f_{N_0,-}$ and $f_{N_0,+}$, taking a given SNR und pulse width Δ_t. As large gap widths, when pulses are separated in time domain, can be resolved by a simple time-of-flight measurement [14], the primary interest lies in the lower limit of w. Inserting (12) into (8) gives

$$w_{min} = \frac{s c_0}{2f_0 + \frac{4\sqrt{\ln(SNR)\ln(2)}}{\pi \Delta_t}}. \tag{13}$$

This provides a closed form estimate for the theoretical detection limit w_{min} of the interlayer width between bone and implant, when using (1) for width calculation.

2.4 Optimal Excitation Frequency

For a plain wave travelling through tissue of thickness l, the intensity reduction is [7]

$$I = I_0\, \mathrm{e}^{-\alpha(f) l} \tag{14}$$

introducing the frequency dependent attenuation factor $\alpha(f)$. This factor usually follows an empirical power law

$$\alpha(f) = \alpha f^q \tag{15}$$

with $q \approx 1$ for most technical and biological materials [7]. Consequently, aiming for the highest signal amplitude, one should reduce the transducer frequency f as far as possible. However, in the considered measurement task of a layered medium with limited thickness of each layer, there is a lower bound for the usable frequency. While pulse overlapping within the soft-tissue interlayer is no problem for measuring the interlayer width due to the spectral minima approach proposed by Lützelberger et al. [8], the limited thickness of the cortical bone layer is a critical issue: As soon as the travel time t_{Bone} within the bone layer gets smaller than the pulse duration, multiple frontside and backside reflections within the bone layer start to overlap and (1) is not applicable anymore for calculating the soft tissue interlayer width. Consequently, there is the criterion

$$t_{Bone} = \frac{2d_{Bone}}{c_{Bone}} > \frac{\beta}{f} \tag{16}$$

with the thickness d_{Bone} of and speed of sound c_{Bone} within the bone layer as well as the number β of oscillations performed by the transducer used.
(16) as a lower frequency bound always has to be fulfilled for enabling correct interlayer width determination.
According to (14), (15) and (16), the theoretically optimal excitation frequency is the lowest frequency enabling time-domain pulse separation within the cortical bone layer, i.e., fulfilling (16). This leads to

$$f_{opt} = \frac{\beta c_{Bone}}{2d_{Bone}} \tag{17}$$

as the optimal excitation frequency.

3 Results

3.1 Minimum Detectable Gap Width

Using realistic parameters from Lützelberger et al. [8]:
- Sound velocity: $c_0 = 1460$ m/s
- FWHM pulse width: $\Delta_t = 1.0$ µs

the theoretical detection limits for selected values of SNR, calculated with (11), become:
- For $SNR = 20000$: $w_{min} = 156$ µm
- For $SNR = 200$: $w_{min} = 173$ µm
- For $SNR = 2$: $w_{min} = 212$ µm

These values match the experimentally observed detection capabilities reported by Lützelberger et al. [8], who achieved measurable gap widths down to ≈ 160 µm under idealized conditions and gap widths down to ≈ 210 µm under more realistic conditions.

3.2 Optimal Excitation Frequency

Regarding the future clinical application, the smallest possible bone layer thickness $d_{Bone,min}$ as well as the largest possible sound velocity $c_{Bone.max}$ must be considered when calculating the optimal excitation frequency by (17). Using typical anatomical values
- Minimal bone thickness $d_{Bone,min} = 3.3$ mm [15]
- Maximal sound speed $c_{Bone.max} = 4600$ m/s [16]

and $\beta \approx 4$ considering the experimental findings by Lützelberger et al. [8], one obtains

$$f_{opt} \approx 2.79 \text{ MHz}.$$

This aligns with the experimental choice and confirms the suitability of the 3 MHz probe used in [8].

4 Discussion

The detection-limit analysis shows that the minimal resolvable gap width depends primarily on:

1. Pulse duration Δ_t, governed by transducer physics.
2. SNR, but only logarithmically – meaning that improving SNR beyond moderate values yields diminishing returns.
3. Sound speed and anatomical conditions, which are fixed and patient dependent.

Importantly, the theoretical minimum gap widths derived here (156 – 212 µm) correspond closely to those experimentally measured values of Lützelberger et al. [8]. This agreement indicates that the interferometric method is operating near the physically achievable limits imposed by acoustic uncertainty relations.

Furthermore, the analysis demonstrates that the 3 MHz excitation frequency in the experiments of Lützelberger et al. [8] is near the analytically derived lower frequency limit governed by the limited bone thickness. It represents a good choice as it reduces frequency-dependent tissue attenuation as far as possible.

The results also highlight the fundamental advantage of frequency-domain analysis over time-of-flight methods, which fail entirely for gaps smaller than the wavelength.

5 Conclusion

This study provides a theoretical foundation for two central design choices in interferometric ultrasound for hip implant loosening diagnostics:

- The minimum detectable gap width is constrained by pulse bandwidth and SNR and lies at approx. 156–212 µm for various measurement conditions.
- The excitation frequency of 3 MHz arises from optimizing the trade-off between attenuation and avoiding pulse interference within the cortical bone layer.

These findings match experimental observations and support the continued development of frequency-domain ultrasound for early loosening detection and implant integration monitoring. They also offer guidance for future transducer and system design, particularly regarding pulse shaping and optimal frequency selection.

Future investigations include a detailed exploration of the interactions between minimum detectable gap width and optimal excitation frequency and investigation of further system design aspects, e.g., the transducer assembly, sound field characteristics and non-perpendicular excitation.

6 Literature

[1] Jennison, T.; MacGregor, A.; Goldberg, A.: Hip arthroplasty practice across the organisation for economic co-operation and development (OECD) over the last decade. The Annals of The Royal College of Surgeons of England. Vol. 105 (2023), no. 7, pp. 645-652.

[2] Apostu, D.; Lucaciu, O.; Berce, C., Lucaciu, D.; Cosma, D.: Current methods of preventing aseptic loosening and improving osseointegration of titanium implants in cementless total hip arthroplasty: a review. Journal of International Medical Research. Vol. 46 (2018), no. 6, pp. 2104-2119.

[3] Wirtz, D.C.; Niethard, F.U.: Ursachen, Diagnostik und Therapie der aseptischen Hüftprothesenlockerung – Eine Standortbestimmung. Zeitschrift für Orthopädie und Unfallchirurgie. Vol. 135 (1997), no. 4, pp. 270-280.

[4] Ruther, C.; Timm, U.; Ewald, H.; Mittelmeier, W.; Bader, R.; Schmelter, R.; Lohrengel, A., Kluess, D.: Current possibilities for detection of loosening of total hip replacements and how intelligent implants could improve diagnostic accuracy. Recent Advances in Arthroplasty. InTech. (2012), pp. 363-386.

[5] Lewin, P.A.: Quo vadis medical ultrasound? Ultrasonics. Vol. 42 (2004), no. 1-9, pp. 1-7.

[6] Powers, J.; Kremkau, F.: Medical ultrasound systems. Interface Focus. Vol. 1 (2011), no. 4, pp. 477-489.

[7] Wells, P.N.T. (Ed.): Ultraschall in der medizinischen Diagnostik. Berlin und Boston: De Gruyter, 2019.

[8] Lützelberger, J.; Arneth, P.; Franck, A.; Drese, K.S.: Ultrasonic Interferometric Procedure for Quantifying the Bone–Implant Interface. Sensors. Vol. 23 (2023), no. 13, pp. 5942.

[9] Puleo, D.: Understanding and controlling the bone–implant interface. Biomaterials. Vol. 20 (1999), no. 23-24, pp. 2311-2321.

[10] Reddyhoff, T.; Kasolang, S.; Dwyer-Joyce, R.S., Drinkwater, B.W. The Phase Shift of an Ultrasonic Pulse at an Oil Layer and Determination of Film Thickness. Proceedings of the Institution of Mechanical Engineers, Part J: Journal of Engineering Tribology. Vol. 219 (2005), no. 6, pp. 387-400.

[11] Pialucha, T.; Guyott, C.C.H.; Cawley, P.: Amplitude spectrum method for the measurement of phase velocity. Ultrasonics. Vol. 27 (1989), no. 5, pp. 270-279.

[12] Kinra, V.K.; Jaminet, P.T.; Zhu, C.; Iyer, V.R.: Simultaneous measurement of the acoustical properties of a thin-layered medium: The inverse problem. The Journal of the Acoustical Society of America. Vol. 95 (1994), no. 6, pp. 3059-3074.

[13] Mallat, S.: A wavelet tour of signal processing. San Diego, London: Academic Press – An Imprint of Elsevier, 1999.

[14] Kinra, V.K.; Iyer, V.R.: Ultrasonic measurement of the thickness, phase velocity, density or attenuation of a thin-viscoelastic plate. Part I: the forward problem. Ultrasonics. Vol. 33 (1995), no. 2, pp. 95-109.

[15] Croker, S.L.; Clement, J.G.; Donlon, D.: A comparison of cortical bone thickness in the femoral midshaft of humans and two non-human mammals. Homo. Vol. 60 (2009), no. 6, pp. 551-565.

[16] Salehi, L.: Nonlinear tomographic reconstruction of elastic properties of isotropic solid materials from ultrasound measurements. Ph.D. Thesis, Ruhr Universität, Bochum, Germany, 2016, p. 5. Available online: http://hss-opus.ub.ruhr-unibochum.de/opus4/frontdoor/index/index/docId/5208.

DOI 10.5162/iCCC2026/7.4

Modellierung einer Heat-Pipe-gestützten Thermalregulierung von HT-PEM-Brennstoffzellen in der Luftfahrt - Numerische Bestimmung geometriespezifischer Wärmeübergangskoeffizienten

Fabian Kramer, Friedrich Franke, Markus Kober, Stefan Kazula
Deutsches Zentrum für Luft- und Raumfahrt e.V. (DLR), Lieberoser Straße 13a, Cottbus, Deutschland
fabian.kramer@dlr.de

Kurzfassung

Zur Verbesserung des Thermalmanagements von Hochtemperatur-Polymerelektrolytmembran-Brennstoffzellen (HT-PEMFCs) wird ein Konzept untersucht, bei dem hexagonähnliche Pulsierende Heat Pipes (PHPs) in die Bipolarplatten integriert werden. Dazu wird ein bestehendes 1D-Modell durch Computational Fluid Dynamics (CFD)-basierte, geometriespezifische Wärmeübergangskoeffizienten erweitert. Die Validierung erfolgt anhand eines kreisförmigen Referenzfalls, der mit Literaturdaten verglichen wird. Für die hexagonähnliche Geometrie zeigt sich, dass das ursprüngliche Modell den Wärmeübergang überschätzt und das aktualisierte Modell eine um etwa 7 K nach oben verschobene Temperaturverteilung vorhersagt. Trotz dieser Abweichung ist das betrachtete Konzept weiterhin realisierbar.

1 Einleitung

Der Luftfahrtsektor steht zunehmend unter Druck, seine Treibhausgasemissionen drastisch zu senken, um einen Beitrag zur Eindämmung der globalen Klimaveränderung zu leisten. Die Europäische Kommission fordert in ihrer Vision Flightpath 2050 technologische Durchbrüche, die eine Reduktion der CO_2-Emissionen um 75 % pro Passagierkilometer im Vergleich zu Flugzeugen des Jahres 2000 ermöglichen [1]. Aktuelle Strategiepapiere und wissenschaftliche Analysen zeigen, dass kein einzelner Technologiepfad die Dekarbonisierung der Luftfahrt bis 2050 allein gewährleisten kann.

Für die Langstrecke gelten nachhaltige Flugkraftstoffe (SAF) als unverzichtbare Übergangslösung, da sie in bestehenden Triebwerken eingesetzt werden können und ein großes Reduktionspotenzial bieten [2–4]. Auf Kurz- und Mittelstrecken rücken ab den 2030er-Jahren hingegen Wasserstoffantriebe in den Fokus, die bei geeigneten Flugzeugkonzepten erhebliche Emissionsminderungen ermöglichen. Ihre Vorteile beruhen vor allem auf der hohen spezifischen Energie von Wasserstoff sowie dem Fehlen von Kohlenstoffemissionen im Betrieb, erfordern jedoch zugleich eine umfassende Infrastruktur und technologische Anpassungen [4–6].

Ein vielversprechender Ansatz zur effizienten Nutzung von Wasserstoff ist der Einsatz brennstoffzellenbasierter elektrischer Antriebssysteme. Polymerelektrolytmembran-Brennstoffzellen (PEMFCs) zählen zu den am weitesten verbreiteten Technologien, da sie eine hohe Effizienz, eine hohe Leistungsdichte und einen weitreichenden Leistungsbereich aufweisen [7]. Sie werden in Low-Temperature- (LT) und High-Temperature-Varianten (HT) unterschieden, basierend auf ihrem jeweiligen Betriebsbereich. HT-PEMFCs, die im Temperaturbereich von 120 bis 200 °C betrieben werden, eignen sich besonders für den Einsatz in der Luftfahrt. Im Vergleich zu LT-PEMFCs zeichnen sie sich durch ein geringeres Systemgewicht, eine vereinfachte thermische Integration und eine größere Leistungsdichte aus [8]. Darüber hinaus begünstigt die höhere Temperatur eine schnellere Reaktionsgeschwindigkeit, Wasser liegt ausschließlich in einem Aggregatzustand vor und der größere Temperaturunterschied zwischen HT-PEMFC und Umgebung erleichtert das Thermalmanagement [9].

Auf Zellebene besteht eine PEMFC aus zwei Bipolarplatten und einer Membran-Elektroden-Anordnung (MEA). Letztere setzt sich aus zwei Gasdiffusionsschichten, zwei Katalysatorschichten sowie einer Polymermembran zusammen [10]. Auf der Anodenseite wird Wasserstoff und auf der Kathodenseite Sauerstoff über Kanäle innerhalb der Bipolarplatten in die Zelle eingeleitet und anschließend über die Gasdiffusionsschichten verteilt. In der Katalysatorschicht wird der Wasserstoff in Protonen und Elektronen aufgespalten. Während die Protonen die Polymermembran passieren, werden die Elektronen über die leitfähigen Bipolarplatten zur Kathodenseite weitergeleitet. Auf diese Weise entsteht ein geschlossener Stromkreis, in dem die Elektronen als nutzbare elektrische Energie abgegriffen werden können. Protonen und Elektronen reagieren schließlich mit Sauerstoff zu Wasser. Dieses wird zusammen mit den unverbrauchten Reaktanden über die Kanäle in den Bipolarplatten wieder aus der Zelle abgeführt.

Neben elektrischer Energie und Wasser entsteht bei der elektrochemischen Reaktion auch ein erheblicher Wärmeanteil. Reale Systeme wandeln nur 40–50 % der chemischen Energie in Elektrizität um [9, 11]. Die restliche Energie fällt als Abwärme an, hauptsächlich in der Kathodenkatalysatorschicht durch ohmsche Verluste und Reaktionswärme [12].

Damit ein stabiler Betrieb gewährleistet ist, muss diese thermische Last durch ein Thermalmanagementsystem (TMS) abgeführt werden. Ziel eines TMS ist es, die Membran im optimalen Temperaturbereich zu halten und eine gleichmäßige Temperaturverteilung über die aktive Zellfläche sicherzustellen [11, 13].

Brennstoffzellen, die dem aktuellen Stand der Technik entsprechen, nutzen in der Regel eine Flüssigkeitskühlung [11]. Dabei wird ein Kühlmittel durch Kanäle in den Bipolarplatten geleitet, erwärmt und anschließend über Wärme-

DOI 10.5162/iCCC2026/7.4

tauscher an die Umgebung abgeführt. Dieses Verfahren ermöglicht eine präzise Temperaturregulierung und eine homogene Belastung der Zellen, erfordert jedoch zahlreiche Zusatzkomponenten wie Pumpen, Leitungen, Ventilatoren und voluminöse Wärmetauscher. Gewicht, Bauraumbedarf und zusätzlicher parasitärer Energieverbrauch stellen insbesondere in der Luftfahrt erhebliche Nachteile dar.
Daher rücken alternative Thermalmanagementstrategien stärker in den Fokus. Grundsätzlich können TMS in vier Kategorien eingeteilt werden. Hierunter zählen konduktive Kühlung, Luftkühlung, Flüssigkeitskühlung und Phasenwechselkühlung [13, 14]. Eine vergleichende Studie dieser Konzepte im Luftfahrtkontext zeigt das Potenzial von Heat-Pipe-basierten Verfahren hinsichtlich thermischer Leistung, Gewicht und Systemzuverlässigkeit [14].
Darauf aufbauend wurde in einer Untersuchung ein Konzept entwickelt, das auf einer passiven konduktiven Kühlmethode basiert [15]. Dabei kommen Pulsating Heat Pipes (PHPs) als alternative Kühlstrategie für HT-PEM-Brennstoffzellen zum Einsatz. PHPs sind geschlossene kapillare Rohrsysteme, die teilweise mit einem Arbeitsfluid gefüllt sind. Durch periodische Verdampfung und Kondensation des Fluids sowie durch kapillare Kräfte transportieren sie Wärme effizient entlang ihrer Längsausdehnung [16].
Im vorgeschlagenen Konzept werden PHPs mit hexagonähnlichem Querschnitt in die Bipolarplatten integriert [15]. Eine schematische Darstellung ist in Bild 1 gezeigt. Das heiße Ende der Struktur steht dabei in direktem Kontakt mit der MEA, in der die Reaktionswärme entsteht, während das kalte Ende über die HT-PEMFC hinausgeführt und in einen externen Kühlluftstrom eingebettet wird.

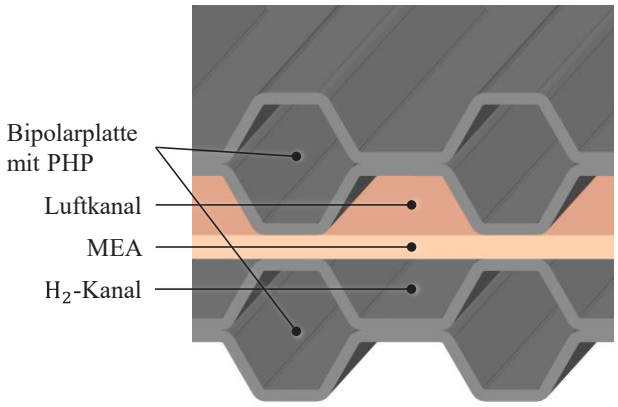


Bild 1 Schematischer Querschnitt der MEA-Baugruppe mit zwei Bipolarplatten und integrierten PHPs [15]

Zur theoretischen Untersuchung dieser Methodik wurde ein eindimensionales Simulationsmodell entwickelt, mit dem verschiedene Eingabeparameter variiert wurden [15]. Die Ergebnisse zeigen, dass die Betriebstemperatur für eine Vielzahl an Parametern unter 200 °C gehalten werden kann. Zudem wird deutlich, dass entlang eines PHP-Strangs ein Temperaturunterschied von maximal 10 K, der als Kriterium für eine homogene Temperaturverteilung dient, nur erreicht wird, wenn die thermische Leitfähigkeit der PHP größer als 3300 W/mK ist.

Trotz vielversprechender Ergebnisse weist das verwendete 1D-Modell Einschränkungen auf. Der Wärmeübergangskoeffizient wird über eine empirische Korrelation bestimmt, deren Gültigkeit für die hexagonähnliche Geometrie bislang nicht ausreichend untersucht wurde. Da dieser Koeffizient maßgeblich die Wärmeabgabe an die Umgebung bestimmt, besteht eine zentrale Unsicherheit.
Vor diesem Hintergrund verfolgt die vorliegende Arbeit das Ziel, das bestehende Modell durch eine zweidimensionale Computational Fluid Dynamics (CFD)-Simulation in Ansys Fluent 2024 R1 zu erweitern, die geometriespezifische Wärmeübergangskoeffizienten liefert.

2 Simulationsmodell

Zur Verbesserung der Vorhersage des konvektiven Wärmeübergangskoeffizienten für den hexagonähnlichen Querschnitt wird ein 2D-CFD-Modell entwickelt. Da für diese Geometrie bislang keine numerischen oder experimentellen Daten vorliegen, wird zunächst ein kreisförmiger Referenzquerschnitt mit einem Durchmesser von 5 mm simuliert und mit Literaturdaten validiert. Anschließend wird die Methodik auf die hexagonähnliche Geometrie mit einem Innendurchmesser von 2 mm und einer Wandstärke von 0,2 mm übertragen.
Insgesamt wird ein Bereich von Reynoldszahlen zwischen 100 und 28.000 untersucht, sodass sowohl laminare als auch turbulente Strömungsbedingungen abgedeckt sind.

2.1 Rechendomain

Die Rechendomain ist als rechteckiger Querschnittskanal ausgelegt und in Bild 2 schematisch dargestellt. Ihre Höhe und Länge werden über dimensionslose Verhältnisse zur horizontalen Ausdehnung der untersuchten Geometrie definiert. Die exakten Werte werden im Rahmen einer Netzunabhängigkeitsstudie bestimmt, sodass äußere Randbedingungen keinen Einfluss auf den Wärmeübergang haben. Das Querschnittsprofil wird als Aussparung in die Rechendomain integriert und befindet sich bei 40 % der Gesamtlänge, gemessen von der linken Seitenwand. Um das Profil wird eine quadratische Verfeinerungszone definiert. Dadurch wird die Gitterauflösung in unmittelbarer Nähe des Profils erhöht, während im Fernfeld eine gröbere Diskretisierung möglich ist. Direkt am Querschnitt wird eine der Geometrie entsprechende Inflationszone erstellt, um die wandnahen Strömungsgrößen adäquat aufzulösen.

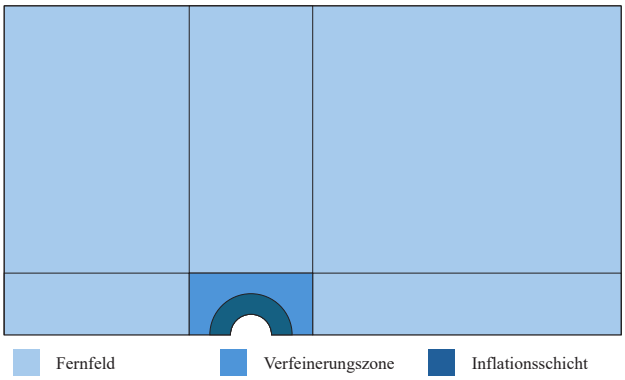


Bild 2 Schematische Darstellung der Rechendomain

DOI 10.5162/iCCC2026/7.4

Zur Reduzierung des Rechenaufwands wird die Domain entlang der horizontalen Symmetrieachse halbiert, da die untersuchten Geometrien symmetrisch aufgebaut sind.

2.2 Modell

Die vertikalen Ränder der Rechendomain dienen als Ein- und Auslass. An der linken Seitenwand wird eine konstante Strömung senkrecht zur Wand mit einer Temperatur von 323,15 K initialisiert. Am Auslass wird eine Druckrandbedingung von 0 Pa angewendet, sodass die Strömung ungehindert austreten kann.
Als Arbeitsmedium dient Luft mit konstanten Stoffwerten. Die Inkompressibilität ist durch eine Machzahl kleiner als 0,3 gewährleistet. In Bezug auf die konvektive Wärmeübertragung wird konservativ nur erzwungene Konvektion berücksichtigt, was bei kleineren Reynoldszahlen zu einer Unterschätzung des Wärmeübergangs führt. Viskose Dissipation ist aktiviert, bleibt jedoch für alle Simulationen vernachlässigbar, da die Brinkmanzahl kleiner als eins ist.
Die obere horizontale Wand wird als stationäre Wand mit No-Slip-Bedingung und adiabatischen Eigenschaften definiert. Die Symmetrieebene durch das Querschnittsprofil wird als Symmetrie-Randbedingung modelliert. Die Wand des Profils wird als No-Slip-Oberfläche mit konstanter Wandtemperatur behandelt. Die Gültigkeit dieser Annahme ist durch eine Knudsenzahl kleiner 0,001 bestätigt.

2.3 Berechnungsverfahren

Die Berechnungen erfolgen mit dem gekoppelten Druck–Geschwindigkeits-Lösungsansatz in ANSYS Fluent 2024 R1. Für die Turbulenzmodellierung wird das SST k–ω-Modell eingesetzt.
Die Konvergenz gilt als erreicht, wenn Residuen und überwachte Größen unter 10^{-6} fallen. Überwacht werden flächenintegrierte Werte entlang der Querschnittswand, darunter Wärmestrom, Wärmeübergangskoeffizient, Nusseltzahl, Temperatur und y^+-Wert. Zusätzlich wird die Luftströmungstemperatur kontrolliert.

2.4 Netzunabhängigkeitsstudie

Die Zuverlässigkeit numerischer Simulationen setzt eine hinreichend feine und zugleich effiziente Diskretisierung voraus. Um sicherzustellen, dass die Ergebnisse nicht von der Gitterauflösung abhängen, wird eine Netzunabhängigkeitsstudie durchgeführt. Da die Geometrie vergleichsweise einfach aufgebaut ist, wird die gesamte Rechendomain in die Untersuchung einbezogen.
Als Ausgangspunkt dient ein gleichförmiges Elementraster, das über die gesamte Geometrie gelegt wird. Darauf aufbauend werden die Dimensionen des rechteckigen Querschnittskanals variiert. Dabei wird die Höhe, die Länge und die Abmaße der quadratischen Verfeinerungszone untersucht. Zur Reduzierung des Rechenaufwands werden zusätzlich Wachstumsraten für Höhe und Länge des Kanals betrachtet. Es zeigt sich, dass eine Variation entlang der Strömungsrichtung möglich ist, während eine gleichzeitige Anpassung der vertikalen Wachstumsrate bei höheren Reynoldszahlen zu fehlender Konvergenz führt.
Ein besonderer Fokus liegt auf der Inflationsschicht an der Profilwand. Um die Wandgradienten ausreichend fein erfassen zu können, werden die Höhe der ersten Schicht, die Wachstumsrate und die Gesamthöhe der Schicht angepasst sowie die Anzahl der Unterteilungen variiert.
Die Untersuchung wird für Reynoldszahlen von 100 und 28.000 sowie Profilwandtemperaturen von 333,15 K und 473,15 K durchgeführt. Die Ergebnisse zeigen, dass die berechneten Werte in Bezug auf die kreisförmige und hexagonähnliche Geometrie netzunabhängig sind, da sämtliche Änderungen der betrachteten Kenngrößen bei weiteren Netzverfeinerungen stets unter 0,37 % liegen.

2.5 Validierung des Modells

Zur Validierung des entwickelten CFD-Modells wird eine Geometrie mit kreisförmigem Querschnitt untersucht. Die Simulationen erfolgen bei einer konstanten Profilwandtemperatur von 473,15 K und werden mit der in dem eindimensionalen Modell verwendeten empirischen Korrelation nach Gnielinski [17] als auch mit den experimentellen Messdaten von Hilpert [18] verglichen. Da die Korrelation nach Gnielinski eine alternative Definition der charakteristischen Länge verwendet, die als Kreisdurchmesser multipliziert mit $\pi/2$ festgelegt ist, ergeben sich für die Reynolds- und Nusseltzahlen abweichende Werte. Um die Vergleichbarkeit sicherzustellen, werden die Simulations- und Versuchsdaten daher entsprechend skaliert. Die Resultate sind in Bild 3 dargestellt.

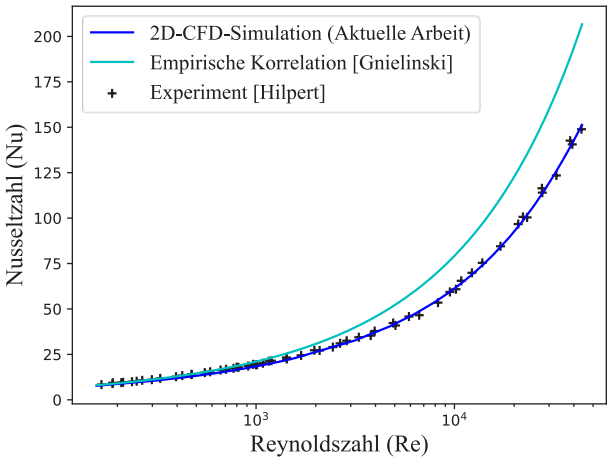


Bild 3 Vergleich der Nusseltzahlen aus der 2D-CFD-Simulation eines kreisförmigen Querschnitts mit empirischer Korrelation [17] und experimentellen Daten [18]

Für Reynoldszahlen kleiner als 1000 zeigen alle drei Datensätze eine enge Übereinstimmung. In diesem Bereich überschätzt die Korrelation nach Gnielinski die Nusseltzahl leicht im Vergleich zu den experimentell gemessenen Werten, während die CFD-Simulation etwas darunterliegt, was die konservative Vorhersage des numerischen Modells bestätigt. Bei Reynoldszahlen größer als 1000 weicht die Korrelation hingegen vom Trend der übrigen Datensätze ab und prognostiziert einen stärkeren Anstieg der Nusseltzahl. Die CFD-Simulationen bleiben dagegen in sehr guter

DOI 10.5162/iCCC2026/7.4

Übereinstimmung mit den Messungen von Hilpert und erfassen das konvektive Verhalten in diesem Bereich präzise. Insgesamt zeigt der Vergleich, dass das CFD-Modell das experimentelle Verhalten bei hohen Profilwandtemperaturen über den gesamten betrachteten Bereich der Reynoldszahlen mit hoher Genauigkeit reproduziert. Auf dieser Grundlage wird die neue Methode zur Bestimmung des Wärmeübergangskoeffizienten in das bestehende eindimensionale Modell integriert, wobei die Ergebnisse der parametrischen CFD-Simulationen herangezogen werden.

3 Resultate

Nach der Validierung des CFD-Ansatzes am kreisförmigen Querschnitt wird die Methodik auf die hexagonähnliche Geometrie angewendet, die im vorgestellten TMS betrachtet wird. Da der Wärmeübergangskoeffizient die zentrale zu verbessernde Kenngröße darstellt, wird er direkt aus den CFD-Simulationen bestimmt. Für den Vergleich wird die bislang verwendete empirische Korrelation nach Gnielinski [17] herangezogen. Die aus der Korrelation berechneten Nusseltzahlen werden in die entsprechenden Wärmeübergangskoeffizienten umgerechnet. Bild 4 zeigt den Vergleich beider Ansätze, wobei die Simulationsergebnisse für eine Profilwandtemperatur von 473,15 K dargestellt sind.

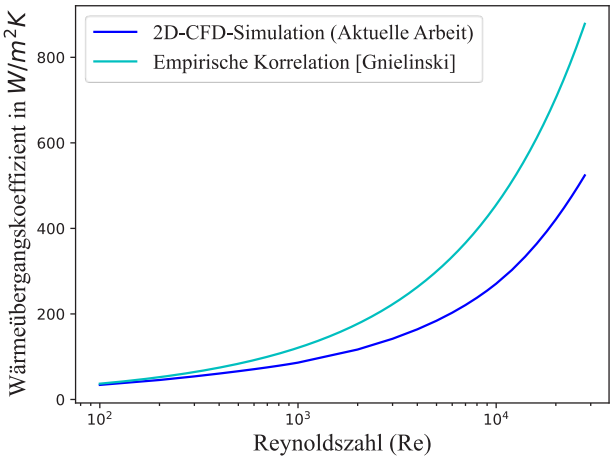


Bild 4 Vergleich des Wärmeübergangskoeffizienten aus der 2D-CFD-Simulation mit empirischer Korrelation [17]

Die Ergebnisse weisen grundsätzlich einen ähnlichen Trend wie im Fall des Kreisquerschnitts (vgl. Bild 3) auf, jedoch treten Abweichungen bereits bei deutlich niedrigeren Reynoldszahlen auf. Nur im Bereich um Re = 100 stimmen die Datensätze gut überein, bei höheren Werten drifften sie zunehmend auseinander. Während die Korrelation nach Gnielinski einen steileren Anstieg prognostiziert, zeigen die CFD-Simulationen ein flacheres Verhalten. Daraus folgt, dass das 1D-Modell auf Basis der Korrelation den Wärmeübergangskoeffizienten überschätzt.

Die Auswirkungen auf die Ergebnisse des eindimensionalen Modells sind in Bild 5 dargestellt. Die Abbildung zeigt die Temperaturverteilung eines PHP-Strangs nach ursprünglicher und aktualisierter Methodik. Das Modell wird in beiden Fällen mit einer Wärmeleitfähigkeit von 3300 W m^{-1} K^{-1} sowie einer Strömung mit einer Geschwindigkeit von 3 m s^{-1} und einer Temperatur von 323,15 K berechnet. Im neuen Modell sind zusätzlich die Punkte markiert, ab denen CFD-basierte Wärmeübergangskoeffizienten verwendet werden.

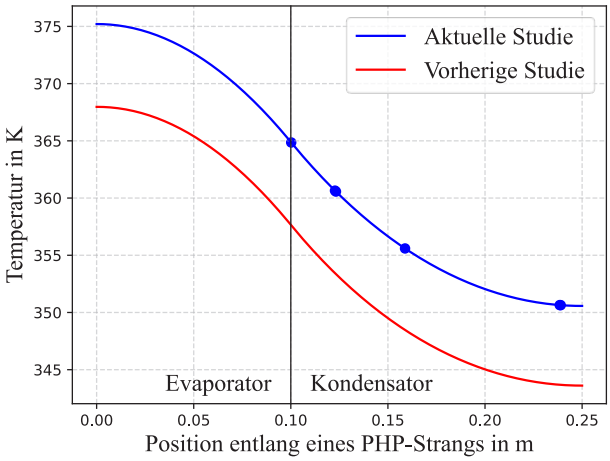


Bild 5 Vergleich der Temperaturverteilungen entlang eines PHP-Strangs zwischen dem ursprünglichen 1D-Modell und der aktualisierten Version

Die beiden Kurven zeigen über den gesamten Bereich hinweg einen übereinstimmenden Verlauf, sodass der Temperaturgradient unverändert bleibt. Gleichzeitig weist das aktualisierte Modell eine um etwa 7 K nach oben verschobene Temperaturverteilung auf.

Insgesamt wird deutlich, dass die CFD-Ergebnisse eine geometriespezifische Wärmeübertragung abbilden und somit eine präzisere Beschreibung des betrachteten Systems ermöglichen. Zudem bestätigt sich, dass die theoretische Umsetzung des vorgestellten Konzepts weiterhin realisierbar ist, da der Maximalwert der Temperaturverteilung unterhalb der Grenzbetriebstemperatur von 473,15 K bleibt.

4 Fazit und Ausblick

Auf Grundlage einer früheren Studie wird ein bestehendes numerisches 1D-Modell erweitert, indem die empirische Korrelation zur Berechnung des Wärmeübergangskoeffizienten durch parametrisierte CFD-Analysen ersetzt wird. Das entwickelte CFD-Modell wird anhand von Literaturdaten für einen kreisförmigen Referenzfall validiert. Die Ergebnisse der hexagonähnlichen Geometrie zeigen, dass der bisherige Ansatz die Wärmeübertragung überschätzt, wodurch die Temperaturen entlang eines PHP-Strangs um rund 7 K unterschätzt werden.

Da für die hexagonähnliche Geometrie bislang keine experimentellen Daten vorliegen, ist eine vollständige Validierung nicht möglich, sodass zur Absicherung und Verfeinerung des Ansatzes weitere Untersuchungen erforderlich sind. In zukünftigen Arbeiten ist hierfür unter anderem der Bau eines Demonstrators des hexagonähnlichen PHP-Kühlsystems vorgesehen. Darüber hinaus sollen künftige Simulationen mehrere PHP-Stränge gleichzeitig berücksichtigen und mit Modifikationen wie Rippen ausstatten, um die Wärmeübertragung zu verbessern. Langfristig soll das Konzept weiterentwickelt werden, um das TMS von Brennstoffzellen zu verbessern und so deren Integration in die Luftfahrt zu unterstützen.

5 Literatur

[1] Directorate-General for Research and Innovation und Directorate-General for Mobility and Transport, *Flightpath 2050: Europe's vision for aviation: maintaining global leadership and serving society's needs* (policy). Luxembourg: Publ. Off. of the Europ. Union, 2011.

[2] E. Cabrera und J. M. M. de Sousa, "Use of Sustainable Fuels in Aviation - A Review," *Energies*, Jg. 15, Nr. 7, 2022, Art. Nr. 2440, doi: 10.3390/en15072440.

[3] Destination 2050, "Roadmap 2025-2050: A Route to Net Zero European Aviation," [Online]. Verfügbar unter: https://www.destination2050.eu/roadmap/

[4] Bundesministerium für Wirtschaft und Energie, "Ergebnisse des Arbeitskreises klimaneutrale Luftfahrt," 2024. [Online]. Verfügbar unter: https://www.bundeswirtschaftsministerium.de/Redaktion/DE/Downloads/klimaschutz/20240604-ergebnisse-des-arbeitskreises-klimaneutrale-luftfahrt.html

[5] Jayant Mukhopadhaya and Dan Rutherford, "Performance analysis of evolutionary hydrogen-powered aircraft," [Online]. Verfügbar unter: https://theicct.org/publication/aviation-global-evo-hydrogen-aircraft-jan22

[6] M. M. Hossain Bhuiyan und Z. Siddique, "Hydrogen as an alternative fuel: A comprehensive review of challenges and opportunities in production, storage, and transportation," *International Journal of Hydrogen Energy*, Jg. 102, S. 1026–1044, 2025, doi: 10.1016/j.ijhydene.2025.01.033.

[7] T. Hoff, F. Becker, A. Dadashi, K. Wicke und G. Wende, "Implementation of Fuel Cells in Aviation from a Maintenance, Repair and Overhaul Perspective," *Aerospace*, Jg. 10, Nr. 1, 2023, Art. Nr. 23, doi: 10.3390/aerospace10010023.

[8] S. Kazula, S. de Graaf und L. Enghardt, "Review of fuel cell technologies and evaluation of their potential and challenges for electrified propulsion systems in commercial aviation," *Journal of the Global Power and Propulsion Society*, Jg. 7, S. 43–57, 2023, doi: 10.33737/jgpps/158036.

[9] Q. Li, R. He, J. O. Jensen und N. J. Bjerrum, "Approaches and Recent Development of Polymer Electrolyte Membranes for Fuel Cells Operating above 100 °C," *Chem. Mater.*, Jg. 15, Nr. 26, S. 4896–4915, 2003, doi: 10.1021/cm0310519.

[10] X.-Z. Yuan und H. Wang, "PEM Fuel Cell Fundamentals," in *PEM Fuel Cell Electrocatalysts and Catalyst Layers*, J. Zhang, Hg., London: Springer London, 2008, S. 1–87.

[11] L. Yang, N.-N. Nik-Ghazali, M. A. Ali, W. T. Chong, Z. Yang und H. Liu, "A review on thermal management in proton exchange membrane fuel cells: Temperature distribution and control," *Renewable and Sustainable Energy Reviews*, Jg. 187, 2023, Art. Nr. 113737, doi: 10.1016/j.rser.2023.113737.

[12] S. G. Kandlikar und Z. Lu, "Fundamental Research Needs in Combined Water and Thermal Management Within a Proton Exchange Membrane Fuel Cell Stack Under Normal and Cold-Start Conditions," *Journal of Fuel Cell Science and Technology*, Jg. 6, Nr. 4, 2009, Art. Nr. 044001, doi: 10.1115/1.3008043.

[13] T. Wilberforce *et al.,* "Recovery of waste heat from proton exchange membrane fuel cells - A review," *International Journal of Hydrogen Energy*, Jg. 52, S. 933–972, 2024, doi: 10.1016/j.ijhydene.2022.08.069.

[14] Fr. Franke, A. Link und S. Kazula, "Evaluation of Heat Transfer Technologies for High Temperature Polymer Electrolyte Membrane Fuel Cells as Primary Power Source in a Regional Aircraft," *in Review bei: CEAS Aeronautical Journal*, 2025.

[15] Fr. Franke, M. Kober und S. Kazula, "Hexagonal Pulsating Heat Pipe for Fuel Cell Integration; Preliminary Design and Evaluation of Thermal Performance," *in Review bei: Case Studies in Thermal Engineering*, 2025.

[16] A. Faghri, "Review and Advances in Heat Pipe Science and Technology," *Journal of Heat Transfer*, Jg. 134, Nr. 12, 2012, Art. Nr. 123001, doi: 10.1115/1.4007407.

[17] V. Gnielinski, "Berechnung mittlerer Wärme- und Stoffübergangskoeffizienten an laminar und turbulent überströmten Einzelkörpern mit Hilfe einer einheitlichen Gleichung," *Forschung im Ingenieurwesen*, Jg. 41, Nr. 5, S. 145–153, 1975, doi: 10.1007/BF02560793.

[18] R. Hilpert, "Wärmeabgabe von geheizten Drähten und Rohren im Luftstrom," *Forschung im Ingenieurwesen*, Jg. 4, Nr. 5, S. 215–224, 1933, doi: 10.1007/BF02719754.

DOI 10.5162/iCCC2026/P1

Analysis of Ultrasound Echoes for Object Properties Characterisation Using Interpretable Features

Priscile Suawa, BTU Cottbus-Senftenberg, Cottbus, Germany, suawapri@b-tu.de
Marcel Jongmanns, Fraunhofer IPMS, Dresden, Germany, marcel.jongmanns@ipms.fraunhofer.de
Christian Herglotz, BTU Cottbus-Senftenberg, Cottbus, Germany, herglchr@b-tu.de

Abstract

This work analyzes ultrasound echoes across multiple objects using a compact set of physically interpretable time- and frequency-domain features, in addition to the raw waveform structure. Unlike black-box models such as deep neural networks or ensembles, which prioritize accuracy over transparency, our approach emphasizes descriptors with direct physical meaning. We extract features that capture geometry, reflectivity, spectral richness, and structural complexity to provide classifiers with interpretable inputs that explain the physical basis for object recognition.
Echoes from five objects were measured across varying orientations and placements. The Time-of-Flight (ToF) decreased near normal incidence and increased at oblique angles, reflecting the propagation geometry. Energy differed by a factor of five between dense-filled and hollow objects, while bandwidth broadened by several hundred hertz under specific orientations. Complementary visualizations using Short-Time Fourier Transform (STFT), Continuous Wavelet Transform (CWT), and Principal Component Analysis (PCA) revealed object-dependent patterns consistent with the scalar features. These results demonstrate that ultrasound echoes encode rich, interpretable information, supporting lightweight feature-based analysis for transparent and low-power embedded sensing.

1 Introduction

Ultrasound sensing is widely used in medical imaging, robotics, and industrial inspection due to its non-invasive nature and sensitivity to both surface and internal structures [1]. While recent machine learning approaches enable object classification from echoes, they often rely on opaque, compute-heavy models with limited physical interpretability, such as deep neural networks [2], [3], which are problematic in safety-critical domains. While these methods can be accurate, they often suffer from high computational complexity and limited interpretability. Attempts to mitigate this, such as GradCAM visualizations [4], provide image-level explanations in diagnostic contexts but do not establish direct links to the physical echo descriptors. Our case study, on the other hand, focuses on directly connecting interpretable features to geometry and material properties.
This work investigates the extent to which information can be directly extracted from echo signals through physics-driven, explainable signal processing. Rather than training deep networks, we extract descriptors such as time-of-flight (ToF), bandwidth, energy, and spectral centroid from ultrasound echoes to characterize material and geometry. These are supplemented with time-frequency visualizations and PCA to reveal structural patterns in the raw signals. Our study focuses on: (1) how angular orientation affects key acoustic features, and (2) how material and structure influence echo signatures across object types. Using five representative objects with varying rigidity, shape, and internal composition at multiple positions, we find that echoes encode rich, interpretable patterns: dense objects yield strong, narrowband reflections, while hollow and porous ones generate resonant responses. These findings highlight that ultrasound echoes contain discriminative and physically meaningful information, eliminating the need for complex computations and providing a foundation for interpretable, low-power embedded sensing.

2 Experimental Setup

Ultrasound echoes were recorded using a custom air-coupled transducer pair centred at 42 kHz, sampled at 2 MHz to preserve acceptable time resolution. The speed of sound in air was 343 m/s (room temperature). Each pulse-echo acquisition produced a 21×6000 matrix, where each row corresponds to an A-scan (Amplitude scan), i.e., a time-domain waveform captured at a specific incidence angle.
Five objects were selected to vary in material (glass, plastic, cardboard), geometry, and fill state (filled vs. hollow). They include a dense glass jar, a hollow plastic flask, a hollow glass bottle, and cardboard boxes of different dimensions. Echoes were collected at three lateral offsets (Left, Middle, Right) and two axial distances (20 cm, 40 cm), with angular variation from −30° to +30°. This resulted in 30 distinct measurement conditions. The test setup ensured consistent placement using a fixed ruler and protractor (see Figure 1). Throughout this paper, we define the axial distance d (object-to-transducer), lateral offset (Left, Middle, Right), and angle α (object rotation) as key geometric parameters.

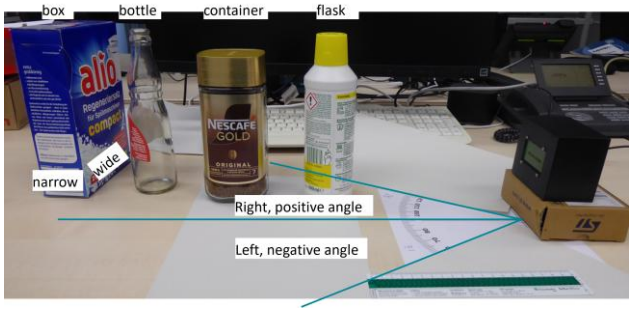


Figure 1. Experimental setup with objects (left) and transducers (right).

3 Methodology

First, the recorded echo measurements undergo pre-processing to remove artifacts and standardize the signal format, before feature extraction. Each A-scan comprises 6000 samples; however, the first 1500 samples are typically dominated by direct coupling and transducer crosstalk. These initial samples are discarded, leaving the remaining 4500 samples for analysis. A zero-phase bandpass filter [5], operating in the range of 35 - 60 kHz, is applied to each signal to isolate the dominant echo frequencies while attenuating low-frequency noise and high-frequency artifacts. This frequency window encompasses the transducer's centre frequency (42 kHz) and its effective bandwidth. Then, to characterize the acoustic responses of different objects across spatial positions and angular orientations, we extract a compact set of physically interpretable features from each A-scan. These include:

- ToF is estimated via peak detection, identifying the first significant reflection as the maximum absolute amplitude in the A-scan [6]. Formally,

$$\text{ToF} = \frac{1}{f_s}\underset{k}{\text{argmax}}\,|x[k]|$$

where $x[k]$ is the discrete-time echo signal, f_s is the sampling frequency (Hz), and k is the sample index of the first significant reflection.

- Echo energy quantifies the total reflected power as the sum of squared amplitudes [7]:

$$E = \sum_{k=0}^{N-1} x\,[k]^2$$

where $x[k]$ is the discrete-time echo signal after preprocessing and N is the number of samples ($N = 4500$ in this work). This corresponds to the time-integrated reflected power, i.e., the signal energy.

- Spectral centroid represents the frequency centre of mass of the echo signal. It is computed from the magnitude spectrum obtained by applying the Fast Fourier Transform (FFT) to the discrete-time signal [8], [9]:

$$C = \frac{\sum_{m=0}^{M-1} f\,[m]\,|X[m]|}{\sum_{m=0}^{M-1}|X[m]|}$$

where $X[m] = \text{FFT}_M\{x[k]\}$ denotes the M-point Fast Fourier Transform of $x[k]$, $|X[m]|$ is its magnitude spectrum, $f[m] = mf_s/M$ is the frequency of bin m, and M is the FFT length (here equal to the number of samples, $N = 4500$).

- Bandwidth quantifies the spectral spread of the signal around its centroid, measuring how concentrated or dispersed the frequency content is:

$$B = \sqrt{\frac{\sum_{m=0}^{M-1}(f[m] - C)^2\ |X[m]|}{\sum_{m=0}^{M-1}|X[m]|}}$$

where $|X[m]|$ is the magnitude spectrum of the signal, $f[m]$ the frequency of bin m, and C the spectral centroid.

Finally, to investigate echo structure beyond scalar features, we apply both the Short-Time Fourier Transform (STFT) and Continuous Wavelet Transform (CWT) [10], [11] to A-scans. The STFT reveals the evolution of spectral energy over time with a fixed spectral resolution, while the CWT provides a multi-resolution analysis, offering finer temporal resolution at high frequencies and higher frequency resolution at low frequencies.

To further explore latent signal structure, PCA [12] is applied directly to the raw A-scan waveforms. PCA projects high-dimensional signals into an orthogonal basis of uncorrelated components, highlighting dominant modes of variation. The leading principal components are retained to reveal object-dependent clustering or geometric separability in feature space.

4 Echo Signature Analysis

In this section, we analysed how ultrasound echoes vary with object geometry, material, and orientation, using scalar features and visualization tools to uncover discriminative patterns.

To evaluate how object orientation influences acoustic signatures, we examine Obj3 (a hollow glass bottle) at a fixed axial distance of 40 cm, with angular sweeps from −30° to +30° in 3° steps across three lateral positions (Left, Middle, Right). We analyse two features: Time-of-Flight (ToF) and Spectral Bandwidth, both sensitive to object geometry and alignment.

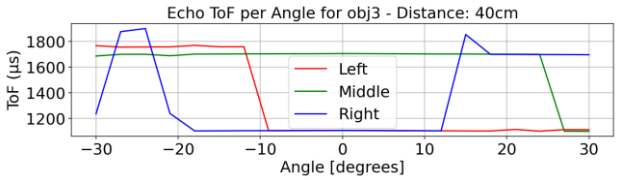


Figure 2. Time-of-Flight (ToF) vs. angle for Obj3 at 40 cm, across three lateral positions.

Time-of-Flight (ToF) measures the delay of the first significant echo. As shown in Figure 2, ToF is shortest near normal incidence ($\alpha \approx 0°$) and increases toward oblique angles due to longer propagation paths. Jagged, step-like variations, especially at off-centre positions, reflect the influence of internal reflections within the hollow structure. Values range from 1100 to 1900 µs, consistent with expectations for a 40 cm round trip in air (c = 343 m/s), after accounting for the discarded 1500-sample pre-echo window. ToF provides a reliable indicator of alignment and echo path length. Variability at oblique angles reveals the role of internal structure, especially in hollow objects like Obj3.

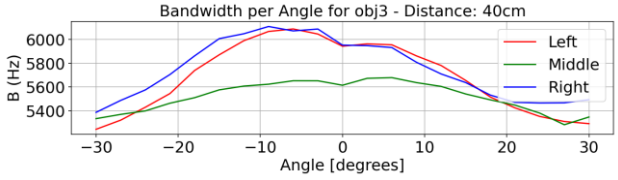


Figure 3. Spectral bandwidth vs. angle for Obj3 at 40 cm.

Spectral Bandwidth, shown in Figure 3, varies more subtly with angle. The Left and Right positions exhibit clear peaks near −10°, reaching up to 5600 Hz, while the Middle position remains relatively flat around 5200 Hz. This suggests that specific orientations enhance high-frequency scattering, possibly due to curvature-induced resonances. Bandwidth, therefore, complements ToF by capturing orientation-dependent spectral richness. Together, these two features reflect complementary aspects of acoustic interaction: ToF traces geometric alignment, while bandwidth encodes frequency-domain effects linked to surface and internal structure. Next, we compared all five objects at a fixed geometry (Middle lateral position, 40 cm axial distance) to assess how material and structure influence echo features. Two key descriptors were selected: energy and spectral centroid.

Echo Energy, as shown in Figure 4, indicates that Obj1 (a filled glass jar) produces the strongest echoes, peaking near 9 × 10^7 a.u. at central angles. This level is nearly four times higher than the hollow glass bottle (Obj3) and plastic flask (Obj2), which reach only 2-2.5×10^7. The cardboard containers (Obj4 narrow box, Obj5 wide box) fall in between, around 4-5×10^7. These contrasts confirm that echo energy is strongly tied to material density and structural fill: dense solids dominate, while hollow or lightweight objects return weaker reflections.

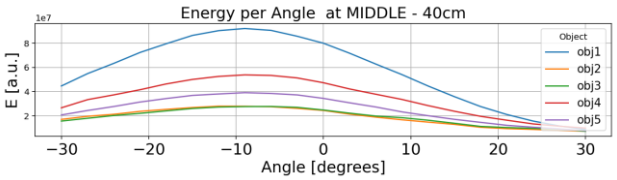


Figure 4. Echo energy (E) in arbitrary units (a.u.) across angle for all objects at Middle position, 40 cm.

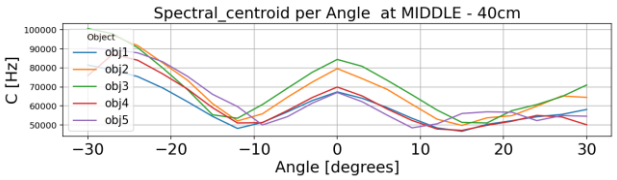


Figure 5. Spectral centroid (C) as a function of angle at Middle position.

Spectral centroid, shown in Figure 5, separates objects by geometry and rigidity. Obj3 reaches up to 100 kHz at oblique angles, approximately 20-30% higher than the cardboard boxes, which remain at around 50-60 kHz. Obj2 also exhibits elevated centroids (up to 80 kHz), reflecting stiff plastic surfaces. These trends suggest that narrow, hollow, or rigid structures concentrate energy in higher frequency bands, while broad, softer materials shift reflections downward.

These two features capture complementary aspects of acoustic behaviour: energy reflects reflectivity and mass, and centroid captures resonance and stiffness. Together, they allow physically interpretable separation of object types.

Table 3 summarizes the physical relevance and observed trends of the retained features, consolidating insights from the prior analyses.

Feature	Physical Link	Observed Trend
Time-of-Flight (ToF)	Propagation geometry	Shortest at normal incidence; increases at oblique angles
Energy	Material density, reflectivity	Highest for dense, filled objects
Spectral Centroid	Resonance, stiffness	Higher for hollow, narrow, or rigid objects
Bandwidth	Spectral richness (angular only)	Peaks at off-normal angles; flatter at centre

Table 3. Summary of features with their physical interpretation and observed trends.

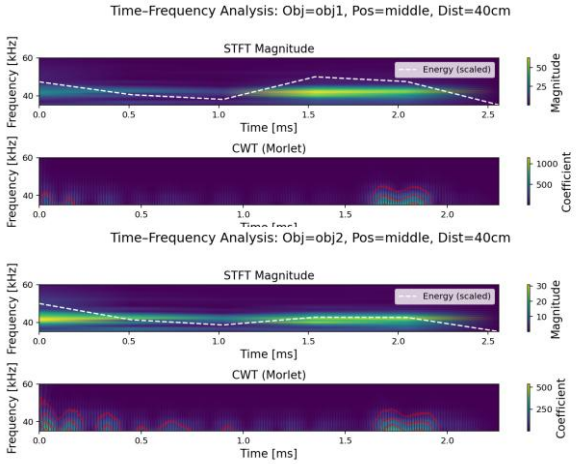


Figure 6. Time-frequency analysis of echoes at d = 40 cm, middle position. STFT magnitude spectrogram with overlaid normalized energy trace (white dashed line). CWT, with red contours indicating regions of high spectral energy.

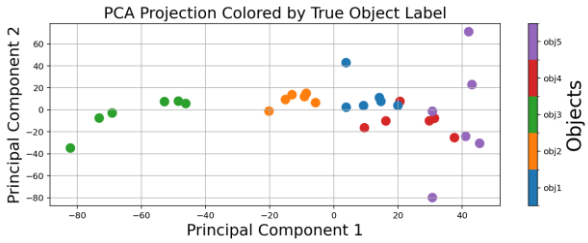


Figure 7. PCA projection of raw waveforms onto the first two principal components.

To complement scalar features, we visualized representative A-scans using time-frequency analysis and dimensionality reduction. STFT and CWT plots (Figure 6) compare Obj1 (filled glass) and Obj2 (hollow plastic). In the STFT panels, Obj1 exhibits a single sharp, short-lived burst of energy visible around 1.5-1.7 *ms*, concentrated in the 38-45 kHz band, consistent with a dense, filled structure producing strong specular reflections. In contrast, Obj2 shows a more prolonged response: its STFT reveals energy distributed over a broader window from 0.5-2.0 *ms*, with frequency content spanning roughly 35-45 kHz, suggesting

reverberation within the hollow cavity. The CWT panels reinforce these differences. For Obj1, the red contours highlight a localized ridge near 40-45 kHz, indicating a compact narrowband reflection. For Obj2, multiple oscillatory ridges persist over time, characteristic of resonant ringing and sustained energy trapping. These complementary views illustrate how hollow and filled objects differ not only in scalar descriptors but also in their temporal and spectral evolution.

Additionally, a PCA projection of raw waveforms (Figure 7) highlights object separability in low-dimensional space. Obj2 and Obj3 form tight, distinct clusters, while Obj1 and Obj4 show partial overlap, indicating that hollow objects produce more consistent and distinctive waveform patterns. Even without explicit feature extraction, raw ultrasound signals exhibit structure that reflects object identity. Time-frequency analysis reveals where energy is concentrated, while PCA confirms waveform-level distinctions.

5 Conclusion and Future Work

This work presents a signal-processing-based framework for ultrasound echo analysis, emphasizing interpretable features directly linked to material, geometry, and structure. Using descriptors such as ToF, energy, spectral centroid, and bandwidth, we demonstrated that even lightweight metrics reveal systematic differences: dense filled objects produce strong, narrowband echoes. At the same time, hollow or porous structures yield weaker and resonant responses.Time-frequency visualizations (STFT, CWT) clarified the origins of scalar features, and PCA revealed separability of object types. Together, these complementary analyses highlight that ultrasound echoes encode rich, object-specific signatures that are both discriminative and physically interpretable.

Future work will first increase the feature set and extend this foundation toward classification by benchmarking raw and engineered features on larger datasets. In parallel, lightweight feature-based classifiers will be compared against opaque black-box models such as CNNs and ensembles, clarifying the trade-off between accuracy and interpretability for low-power embedded sensing.

Such lightweight, interpretable ultrasound characterization could support applications in embedded robotics, packaging inspection, and smart environments, where distinguishing between object types (e.g., filled vs. hollow containers, rigid vs. porous surfaces) must be achieved efficiently under power and computational constraints.

6 Literature

[1] Alexander L. Bowler, Michael P. Pound, Nicholas J. Watson, A review of ultrasonic sensing and machine learning methods to monitor industrial processes, Ultrasonics, Volume 124, 2022, 106776, ISSN 0041-624X, https://doi.org/10.1016/j.ultras.2022.106776.

[2] Shafiei Alavijeh M, Scott R, Seviaryn F, Maev RG. Using machine learning to automate ultrasound-based classification of butt-fused joints in medium-density polyethylene gas pipes. J Acoust Soc Am. 2021 Jul;150(1):561. doi: 10.1121/10.0005656. PMID: 34340509.

[3] Ning Mao, Amirah Nabilah Azman, Guangxin Ding, Yubo Jin, Can Kang, Hyoung-Bum Kim, Black-box real-time identification of sub-regime of gas-liquid flow using Ultrasound Doppler Velocimetry with deep learning, Energy, Vol 239, Part D, 2022, 122319, https://doi.org/10.1016/j.energy.2021.122319.

[4] R. Singh, A. S. Bansal, and M. R. Bansal, "Plaque classification using ultrasound images with machine learning and explainable AI," *Biomedical Signal Processing and Control*, vol. 93, p. 106063, 2024. https://doi.org/10.1016/j.bspc.2023.106063.

[5] Steven W. Smith, The Scientist and Engineer's Guide to Digital Signal Processing, 1st ed. California Technical Publishing, 1997. [Online]. Available: https://www.dspguide.com/ch19/1.htm.

[6] Cherif Othmani, Najmeh Sadat Dokhanchi, Sebastian Merchel, Albert Vogel, M. Ercan Altinsoy, Conrad Voelker, A review of the state-of-the-art approaches in detecting time-of-flight in room impulse responses, Sensors and Actuators A: Physical, Volume 374, 2024, 115467, ISSN 0924-4247, https://doi.org/10.1016/j.sna.2024.115467.

[7] Bowler, A.; Escrig, J.; Pound, M.; Watson, N. Predicting Alcohol Concentration during Beer Fermentation Using Ultrasonic Measurements and Machine Learning. Fermentation 2021, 7, 34. https://doi.org/10.3390/fermentation7010034.

[8] Nilesh Kulkarni, Vinayak Bairagi, Chapter Four - Use of Complexity Features for Diagnosis of Alzheimer Disease, Editor(s): Nilesh Kulkarni, Vinayak Bairagi, EEG-Based Diagnosis of Alzheimer Disease, Academic Press, 2018, Pages 47-59, ISBN 9780128153925, https://doi.org/10.1016/B978-0-12-815392-5.00004-6.

[9] Lv, D.; Zi, J.; Huang, X.; Gao, M.; Xi, R.; Li, W.; Wang, Z. Feature Extraction on the Difference of Plant Stem Structure Based on Ultrasound Energy. Agriculture 2023, 13, 52. https://doi.org/10.3390/agriculture13010052.

[10] Yang Yang, Zhike Peng, Wenming Zhang, Guang Meng, Parameterised time-frequency analysis methods and their engineering applications: A review of recent advances, Mechanical Systems and Signal Processing, Volume 119, 2019, Pages 182-221, ISSN 0888-3270, https://doi.org/10.1016/j.ymssp.2018.07.039.

[11] Lupea, I.; Lupea, M. Continuous Wavelet Transform and CNN for Fault Detection in a Helical Gearbox. Appl. Sci. 2025, 15, 950. https://doi.org/10.3390/app15020950.

[12] Azim, M. R., & Gül, M. (2020). Data-driven damage identification technique for steel truss railroad bridges utilizing principal component analysis of strain response. Structure and Infrastructure Engineering, 17(8),1019–1035. https://doi.org/10.1080/15732479.2020.1785512.

DOI 10.5162/iCCC2026/P2

Physikalisch basiertes virtuelles Prototyping von Mikrosystemen am Beispiel akustischer Wandlersysteme

Physics-based Virtual Prototyping of Microsystems Exemplified for Acoustic Transducer Systems

Gabriele Schrag, Professur für Mikrosensorik und -aktorik, Technische Universität München, München, Deutschland
schrag@tum.de

Kurzfassung

Die große Vielfalt und Komplexität von Sensoren, Aktoren und Wandlern, sowie von den mit ihnen realisierten Mikrosystemen, macht ihre Modellierung und damit ihren Entwurf komplex und herausfordernd. Dieser Beitrag gibt einen Überblick über Herausforderungen von und Anforderungen an physikalisch basierte Modellierungsumgebungen und zeigt Perspektiven für einen ganzheitlichen Entwurf von Mikrosystemen auf. Die demonstrierten Ansätze (Mixed-level Modellierung, Systemsimulation und Optimierung mittels genetischer Algorithmen) sind auf ein breites Spektrum von Bauelementen anwendbar, werden aber exemplifiziert anhand verschiedener akustischer Wandlersysteme.

Abstract

The wide and complex variety of sensors, actuators, and transducers, as well as of microsystems realized with them, makes their modeling and thus their design complex and challenging. This article provides an overview of the challenges and requirements of physics-based modeling environments and outlines perspectives for a holistic design of microsystems. The approaches demonstrated (mixed-level modeling, system simulation, and optimization using genetic algorithms) are applicable to a wide range of components, but are exemplified using various acoustic transducer systems.

1 Virtuelles Prototyping von Mikrosystemen

Die große Vielfalt und Komplexität von Sensoren, Aktoren und Wandlern, sowie von den mit ihnen realisierten Mikrosystemen, macht ihre Modellierung und damit ihren Entwurf komplex und herausfordernd. Insbesondere Aspekte wie die inhärent mit ihrer Wirkweise einhergehenden Kopplungen zwischen verschiedenen Energieformen, die geometrische Komplexität der Mikrostrukturen sowie die Tatsache, dass die Performanz eines Mikrosystems durch alle Teilsysteme (Wandler, Elektronik, Gehäuse, etc.) bestimmt wird, sind hier zu nennen.
Daher werden zuverlässige Entwurfswerkzeuge auf allen Hierarchieebenen des Entwicklungsprozesses benötigt, die modular ineinandergreifen und eine effiziente Simulation erlauben. Idealerweise bleiben dabei physikalische und geometrische Designparameter sowie deren Abhängigkeiten im gesamten Prozess so detailliert wie nötig zugänglich, um den Entwurf solcher Systeme auf Basis einer prädiktiven und physikalisch basierten Modellierung zu ermöglichen. Das Portfolio der über den Designprozess benötigen Tools auf den verschiedenen Hierarchieebenen ist im Überblick in **Bild 1** dargestellt.
Essenziell ist hierbei, dass für alle Designaufgaben adäquate Entwurfswerkzeuge zur Verfügung stehen. Insbesondere sind Methoden notwendig, die die effiziente Modellierung auf Systemebene ermöglichen (siehe roter Rahmen in **Bild 1**) [1].

Im Gegensatz zu elektronischen Systemen, wo standardisierte Entwicklungsplattformen existieren, sind diese aufgrund der Diversität der Wandlerkonzepte bei Mikrosystemen kaum durchgängig vorhanden und oft werden für die jeweiligen Anwendungen gezielt Modelle abgeleitet.

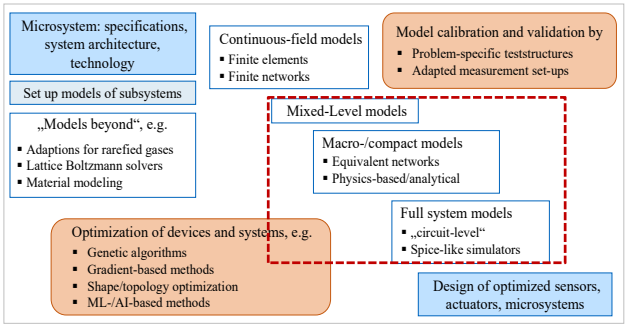


Bild 1 Hierarchische Modellierungsumgebung für das virtuelle Prototyping von Mikrosystemen.

Einen theoretischen Rahmen gibt in diesem Zusammenhang die auf thermodynamischen Prinzipien und Erhaltungsgleichungen [2,3] beruhende sog. verallgemeinerte Kirchhoffsche Netzwerktheorie [4,5], die eine Systemabstrahierung vom verteilten Modell bis zum Systemmodell erlaubt. Die hiermit abgeleiteten Modelle sind voll kompatibel mit gängigen Systemsimulatoren und bieten daher den natürlichen Link zur Mikroelektronik und in die Systemsimulation. Eine dezidiert physikalisch basierte Ableitung von Teilsystemmodellen, die modular zum Gesamtsystem verbunden werden können, ermöglicht dann die gezielte Optimierung von Mikrosystemen auf physikalischer Basis

DOI 10.5162/iCCC2026/P2

und damit eine zuverlässige Ableitung von Designparametern.
Am Beispiel mehrerer akustischer Wandlersysteme werden verschiedene Ansätze zur Modellierung gezeigt. Im Folgenden wird beispielhaft die Optimierung des Ausgangssignals eines Ultraschallwandlers mittels akustischer Hörner durch physikalische Kompaktmodellierung in Verbindung mit genetischen Algorithmen skizziert; weitere Beispiele (Systemsimulation von kapazitiven Mikrofonen mittels Mixed-level-Modellierung [6] sowie die Optimierung piezoelektrischer Mikrofone mittels genetischer Algorithmen[7]) werden detailliert im Konferenzbeitrag adressiert.

2 Beispiel – Maßgeschneiderte Frequenzantwort mit Hilfe akustischer Hörner

2.1 Das System - Problembeschreibung

Die Frequenzantwort von akustischen und Ultraschallwandlern ist stark vom umgebenden Medium und vom akustischen Aufbau (z.B. Montage auf einem PCB) abhängig. Dies kann man sich zunutze machen, indem man akustische Hörner an einen Wandler anschließen lässt und die Form so gestaltet, dass dieser als Gesamtsystem ein bestimmtes Spektrum abstrahlt oder sogar einen bestimmte Direktivität aufweist [8].

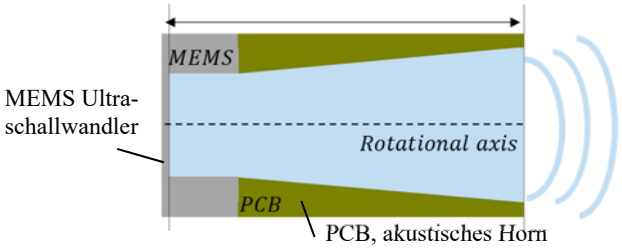


Bild 2 Schematische Darstellung einer Anordnung aus Ultraschallwandler und akustischem Horn

Bild 2 zeigt eine exemplarische Anordnung eine mikroelektromechanischen (MEMS), membranbasierten Ultraschallwandlers in Kombination mit einem akustischen Horn (PCB).

Um das System auf die gewünschte Abstrahlcharakteristik zu trimmen und zu optimieren, sind Modelle nötig, die zum einen effizient genug sind, um Wandler plus Gehäuse und umgebendes Medium gekoppelt zu simulieren, die andererseits aber auch die essentiellen Designparameter (Geometrie, Material) auf Systemebene für eine Optimierung zugänglich hat. Wir verwenden hier Kompaktmodelle für den Wandler wie auch für das akustische Horn, die mittels einer äquivalenten Netzwerkmodellierung oder direkt mittels einer Hardware-Beschreibungssprache (VHLD-AMS, Verilog-A) in einen Standardsystemsimulator implementiert werden können und im folgenden Abschnitt kurz skizziert werden.

2.2 Modellierung und Systemoptimierung

Der prinzipielle Simulationsablauf für die Modellierung und Optimierung des Systems hinsichtlich einer optimierten Horngeometrie für eine gewünschte Abstrahlcharakteristik ist in **Bild 3** dargestellt.
Den Startpunkt stellt ein physikalisch basiertes Kompaktmodell für den elektromechanischen Ultraschallwandler auf Basis einer verallgemeinerten Kirchhoffschen Netzwerkmodellierung dar. Dieses Modell koppelt an eine akustisches Transmission-Line-Modell (TML), das die Ausbreitung der Schallwellen modelliert. Um einen veränderlichen Hornradius in das Modell einbeziehen zu können, wird es in Form eines diskretisierten TLMs verwendet, d.h. es werden mehrere Teilmodelle mit variierenden Querschnittsradien und Längen so aneinandergereiht, das der veränderliche Horndurchmesser emuliert wird (siehe **Bild 4**).
Die zu optimierende Geometrie des akustischen Kanals wird über eine Freiformmodellierung mittels Bezierkurven dargestellt. Durch Anwendung eines genetischen Algorithmus wird die Form des Kanals so optimiert, dass die gewünschte Abstrahlcharakteristik erzielt wird.

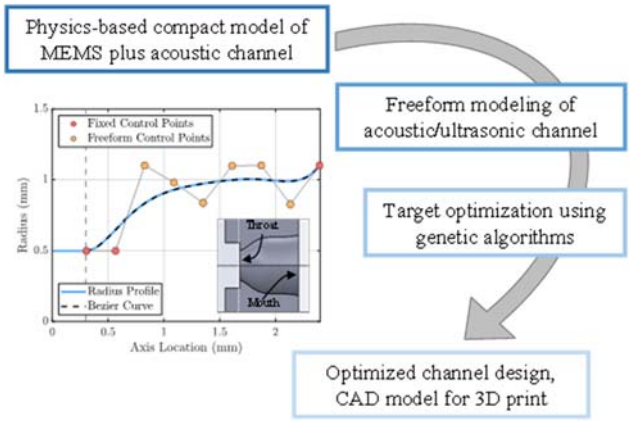


Bild 3 Ablaufschema für die Modellierung und Optimierung des Systems aus mikromechanischem Ultraschallwandler und akustischem Horn.

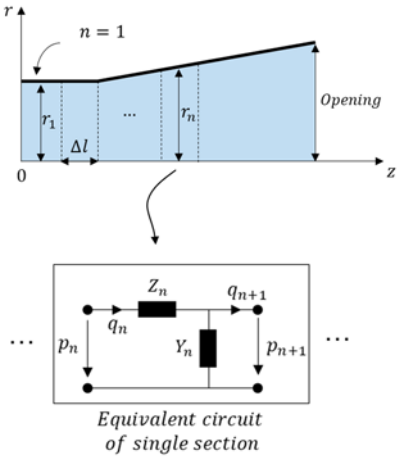


Bild 4 Modellierung des akustischen Horns, mittels eines diskretisierten akustischen Transmissionline-Modells.

DOI 10.5162/iCCC2026/P2

2.3 Simulations- und Messergebnisse

Ausgangspunkt der Untersuchungen war ein Ultraschallwandler mit zylindrischem akustischem Kanal, der mit Hilfe eines Laser-Dopplervibrometers (**Bild 5**, links), das mit einer Vakuumkammer betrieben werden kann, vermessen wurde. In **Bild 6** (rechts) sieht man die Ergebnisse einer Messung im Vakuum (blau), die die mechanische Eigenfrequenz des Wandlers zeigt. Der rot dargestellte Frequenzgang zeigt die Messungen in Luft und die Veränderungen in der Abstrahlcharakteristik, die durch die Umgebungsluft und das akustische Horn verursacht werden.

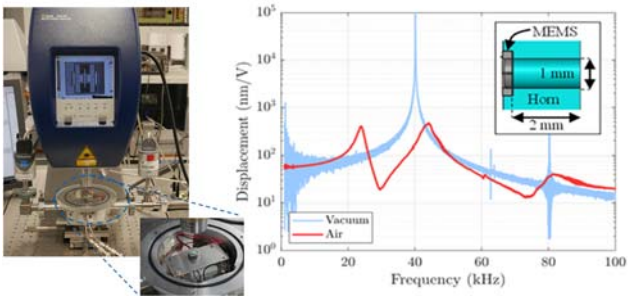


Bild 5 Laser-Doppler-Vibrometer MSA-500 (Polytec) mit Vakuumkammer zur Vermessung der Ultraschallwandler und Frequenzgang der Konfiguration mit zylindrischem Horn im Vakuum (blau) und in Luft (rot).

Mit Hilfe des in 2.2. skizzierten physikalisch basierten Systemmodells und des genetischen Algorithmus wurde die Form des akustischen Horns so optimiert, dass die Abstrahlcharakteristik des Systems auf eine Zielfunktion getrimmt werden konnte und nun Resonanzen bei den vorgegebenen Frequenzen 29 kHz und 41 kHz aufweist.
Die so berechnete Form des akustischen Horns wurde mit Hilfe eines 3D-Druckers realisiert, und das System auf einem PCB aufgebaut und vermessen (siehe **Bild 6**). Der optimierte Frequenzgang ist in **Bild 7** zu sehen und zeigt sehr gute Übereinstimmung zwischen vorhergesagter (optimierter) und gemessener Charakteristik.

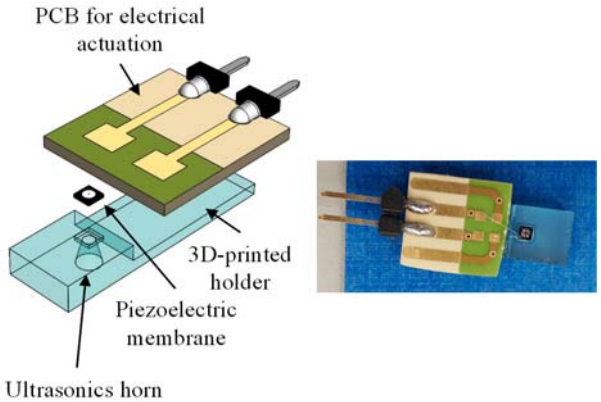


Bild 6 Schematische Darstellung des Aufbaus aus piezoelektrischem Ultraschallwandler, akustischem Horn und PBC (links) und Realisierung (rechts).

3 Zusammenfassung

Die Modellierung mikrotechnologisch gefertigter Sensor- und Aktorsysteme erfordert aufgrund der Diversität und Komplexität der Wandlerprinzipien und Materialsysteme sowie der Komplexität der Bauelementgeometrien eine Entwurfsumgebung, die für alle Abstraktionsebenen Entwurfswerkzeuge bereitstellt, die hierarchisch und modular ineinandergreifen. Da Mikrosysteme mehr als die Summe ihrer Einzelbauteile sind, werden auch Methoden benötigt, die die kompakte, effiziente, aber dennoch akkurate Modellierung auf Systemebene erlauben, wobei alle Effekte, die das Betriebsverhalten beeinflussen, selbstkonsistent und mit allen Rückkopplungen in einem solchen Systemmodell enthalten sein müssen.
Anhand eines Demonstrators aus dem Bereich akustischer Wandler wurde beispielhaft das Potential von, auf physikalischer Modellierung basiertem, virtuellem Prototyping gezeigt. Das Systemmodell, abgeleitet auf Basis der verallgemeinerten Kirchhoffschen Netzwerktheorie, ermöglicht den Zugang zu Designparametern und damit deren effiziente Optimierung auf Systemeben.
Dieses Konzept ist generisch, verallgemeinerbar und daher übertragbar auf andere Mikrosysteme (z.B. [6,7,9]) und bildet daher ein Fundament für den Entwurf derartiger Systeme als Ganzes.

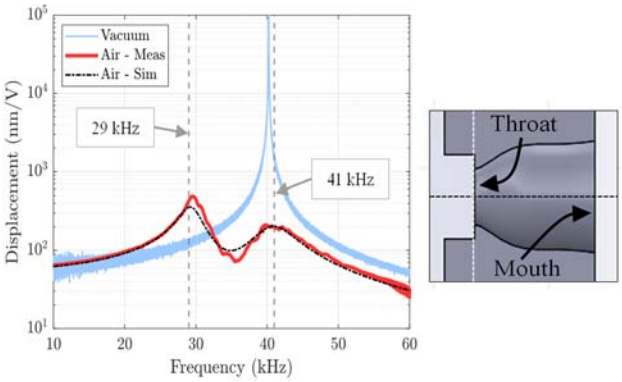


Bild 7 Frequenzgang des optimierten Systems: Vergleich zwischen Simulation und Messung (links) und die durch Optimierung erhaltene Horngeometrie (rechts).

4 Literatur

[1] T. Bechtold, G. Schrag, L. Feng (eds.), System-Level Modeling of MEMS, Advanced Micro & Nanosystems 10. Wiley-VHC, 2013

[2] L. Onsager, “Reciprocal Relations in Irreversible Processes”, Physical Review, vol. 37, 1931.

[3] H.B. Callen, Thermodynamics and an Introduction to Thermostatistics, John Wiley & Sons, New York, 1985

[4] G. Schrag, G. Wachutka, System-Level Modeling of MEMS Using Generalized Kirchhoffian Networks-Basic Principles. In: System-Level Modeling of MEMS, Wiley-VHC, 2013

[5] A. Lenk, R.G. Ballas, R. Werthschuetzky, G. Pfeifer, Electromechanical Systems in Microtechnology and

Mechatronics: Electrical, Mechanical and Acoustic Networks, their Interactions and Applications (Microtechnology and MEMS), Springer, 2012

[6] T. Kuenzig, A. Dehe, U. Krumbein, G. Schrag, Virtual design and optimization studies for industrial silicon microphones applying tailored system-level modeling. Journal of Micromechanics and Microengineering 28, 2018

[7] G. Bosetti, C. Bretthauer, A. Bogner, M. Krenzer, K. Gierl, H. Heiss, G. Schrag, Design Optimization of a Corrugated Piezoelectric MEMS Microphone applying Genetic Algorithms. 2024 25th International Conference on Thermal, Mechanical and Multi-Physics Simulation and Experiments in Microelectronics and Microsystems (EuroSimE), 2024

[8] G. Bosetti, S. Hofstetter-Spona, G. Schrag. On the Use of 3D-Printed Ultrasonic Horns to Tune the Frequency Response of Airborne MEMS Transducers, Transducers Conference, Kyoto, Japan, 2023

[9] P. Voigt, Compact modeling of microsystems, in Selected Topics of Electronics and Micromechatronics (eds G. Wachutka and D. Schmitt-Landsiedel), Shaker Verlag, Aachen, Germany, 2003

DOI 10.5162/iCCC2026/P3

Ultrasonic Waveguide for Multiphase Level Monitoring

Miriam Piechaczek, Alexander Backer, Tobias Roß, Sabrina Tietze, Klaus Stefan Drese,
Institute of Sensor und Actuator Technology / University of Applied Sciences and Arts Coburg, Coburg, Germany
miriam.piechaczek@hs-coburg.de

Abstract

In process industries, continuous and accurate level measurement is essential for safe and efficient operation. However, many industrial fluids are inherently heterogeneous, typically forming stratified layers in which solid particles settle at the bottom while lighter components accumulate near the top. Such multiphase conditions complicate reliable level detection, often requiring multiple sensors to quantify the individual layers. This work presents an alternative approach using a single ultrasonic waveguide to detect the overall fill level and liquid layers by exciting multiple acoustic modes. Each mode exhibits a distinct sensitivity to the surrounding medium, enabling the differentiation of layered structures and improving measurement robustness through temperature compensation. Initial investigations in oil–water mixtures show promising agreement between theoretical predictions and experimental results for two of the three targeted wave modes, demonstrating the feasibility of multimodal guided-wave sensing for complex multiphase environments.

1 Introduction

Continuous and reliable liquid level measurement is a fundamental requirement in numerous applications of the process industry, including wastewater treatment, chemical processing, and oil–water separation. Conventional measurement principles such as capacitive, hydrostatic, or radarbased sensors typically assume homogeneous fluid properties. As a consequence, their accuracy is significantly reduced or the sensors fail entirely when applied to heterogeneous or stratified media [1]. Numerous studies have demonstrated that industrial fluids frequently exhibit pronounced phase separation, with solid particles settling at the bottom of the container while lighter phases accumulate near the surface. These conditions result in complex multiphase environments that pose substantial challenges for conventional level measurement technologies.

In addition to multiphase effects, practical limitations such as sensitivity to foam formation, mechanical fouling, and limited suitability for aggressive or corrosive media further restrict the applicability of conventional level sensors in industrial environments [1]. As a result, multiple sensors or spatially distributed measurement systems are often required to determine both the global fill level and internal phase boundaries. This increases system complexity, installation effort, and maintenance requirements.

Guided acoustic waves have emerged as a promising approach for structural and process monitoring due to their ability to propagate over long distances and their high sensitivity to changes in the surrounding medium. Previous investigations have shown that different acoustic wave modes, including longitudinal, flexural, and torsional modes, interact in characteristic ways with adjacent fluids and boundary conditions. This behavior enables applications ranging from immersion depth estimation to the detection of material discontinuities and fluid interfaces [2–4]. Most earlier studies focused on metallic waveguides such as aluminum or steel rods, where fill level information was derived from attenuation effects or time of flight measurements of specific longitudinal modes [2]. However, the susceptibility of metallic materials to corrosion and their comparatively high material costs limit their applicability in chemically aggressive environments.

In this context, polymer-based waveguides, particularly those made of polyethylene, have gained increasing attention due to their chemical resistance, low material cost, and suitability for the excitation of guided acoustic waves [3,4]. Building on these developments, the multimodal excitation of guided waves within a single polymer waveguide offers the potential to extract layer-dependent information without requiring multiple sensor elements.

In particular, previous studies have shown that torsional guided wave modes exhibit only weak sensitivity to external fluid loading while remaining responsive to temperature dependent material property variations, which makes them suitable as reference modes for temperature compensation in multimodal sensing concepts [5].

In this work, a polymer based ultrasonic guided wave sensor is investigated that exploits multimodal wave–fluid interactions for robust liquid level measurement in heterogeneous fluids. By selectively exciting acoustic modes with different sensitivities to fluid loading, the proposed approach enables the determination of both the global fill level and internal stratification using a single compact sensing element. Furthermore, the integration of fluid sensitive and fluid insensitive guided wave modes establishes the basis for intrinsic temperature compensation within the same sensor system, reducing the need for additional reference sensors and enabling robust operation in chemically aggressive industrial environments.

2 Methods

To investigate the mode specific characteristics with respect to liquid fill level, internal phase boundaries, and temperature dependence, the fill level measurement was validated experimentally and by numerical simulation for longitudinal and flexural guided wave modes. In addition,

DOI 10.5162/iCCC2026/P3

the capability of detecting interfacial layer thicknesses was analyzed by means of simulations. This combined experimental and numerical approach allows the individual influence of geometric, material, and environmental parameters to be isolated and systematically evaluated.
In the experimental setup, excitation signals consisting of ten cycle bursts windowed by a Hanning function are generated using a function generator and applied to a piezoelectric transducer mounted at the top end of the waveguide. The excitation frequency is set to 6 kHz, and the same excitation parameters are used consistently in both the experimental measurements and the numerical simulations. The reflected signals of the L(0,1) and F(1,1) guided wave modes are recorded using an oscilloscope for varying liquid fill levels. As a reference measurement for the liquid height, a hydrostatic pressure sensor is integrated at the bottom of the test container and read out by a digital multimeter connected to the data acquisition and control unit (Figure 1).

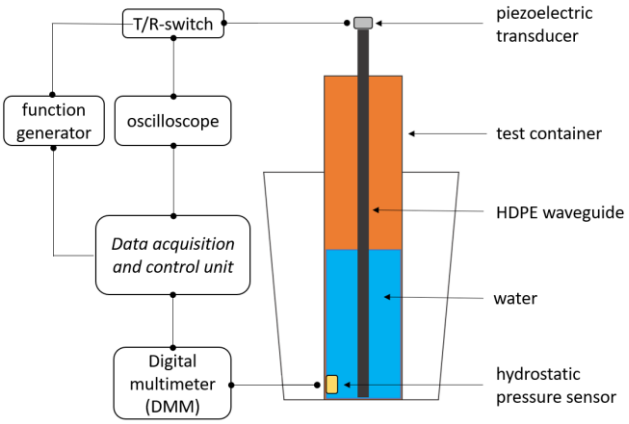


Figure 1: Experimental setup.

The waveguide used in the experiments consists of a high-density polyethylene rod (HDPE) that is vertically immersed in the test container. The choice of a polymer waveguide ensures chemical robustness and enables controlled propagation of multiple guided wave modes. The surrounding medium along the waveguide varies with the liquid fill level, resulting in distinct air coupled and liquid coupled propagation sections.

The selected excitation and evaluation strategy corresponds to a single sensor configuration in which multiple guided wave modes are generated and analyzed using the same waveguide and the same transducer location. This configuration ensures that all observed differences in the measured signals originate from mode specific interactions with the surrounding fluid rather than from spatially separated sensor elements.
The numerical investigations were carried out using a self-developed simulation framework for guided acoustic waves, which has been validated and successfully applied in previous studies [6]. The simulation tool enables the modeling of wave propagation, dispersion effects, and reflections and was adapted to the geometry and boundary conditions considered in this work.

The excitation was applied to a high density polyethylene waveguide with a total length of 1.93 m and a diameter of 40 mm. The waveguide was partially immersed in water at varying fill levels in order to account for relevant fluid structure interactions along the propagation path. The geometry of the waveguide is based on commercially available dimensions and was adopted from previous investigations [1]. The material parameters of the HDPE waveguide were initially taken from manufacturer data sheets. The effective elastic modulus used in the model was experimentally determined [7].
In addition, dispersion curves for the L(0,1) and F(1,1) modes were calculated with the software Disperse for vacuum loaded, water loaded, and oil loaded boundary conditions. This approach allows the influence of external fluid loading on the propagation characteristics of each guided wave mode to be quantified and provides the basis for the mode specific interpretation of experimental time of flight and attenuation effects.

3 Results and Discussion

3.1 Evaluation of liquid fill level

The comparison between numerical simulations and experimental measurements demonstrates that the theoretically predicted wave propagation behavior of the L(0,1) and F(1,1) guided wave modes shows very good agreement with the measured signals over the investigated range of liquid fill levels. The comparison of experiment and simulation is exemplarily shown for the L(0,1) mode in Figure 2.

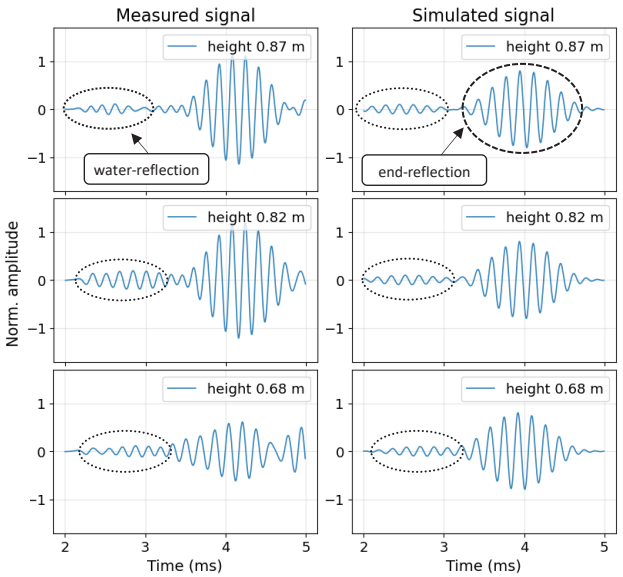


Figure 2: Measured (left) and simulated (right) L(0,1) guided wave signals for varying liquid fill levels.

Both the experimental and simulated signals exhibit a consistent shift in the arrival time of the guided wave groups with varying fill level. In addition to the L(0,1) mode, the F(1,1) mode also shows predictable changes in propagation time, confirming its suitability for the detection of the global liquid fill level. In accordance with previous studies,

the F(1,1) mode exhibits a slightly increased sensitivity to the presence of internal phase boundaries compared to the L(0,1) mode, which makes it particularly attractive for the investigation of stratified fluids [5]. Nevertheless, the evaluation of the L(0,1) mode also shows a robust response to fill level variations. The consistent behavior of both modes enables their use as mutual validation for the determination of the overall liquid height.
The very good agreement between numerical predictions and experimental results for both guided wave modes confirms the validity of the simulation approach. This validated simulation framework is therefore used in the following to investigate mode specific effects in more complex multiphase environments, particularly with respect to the identification of internal phase boundaries.
The signal evaluation is performed using classical zero crossing tracking. In this approach, both the time of flight of the water reflected signal and that of the end reflected signal are considered (Figure 3).

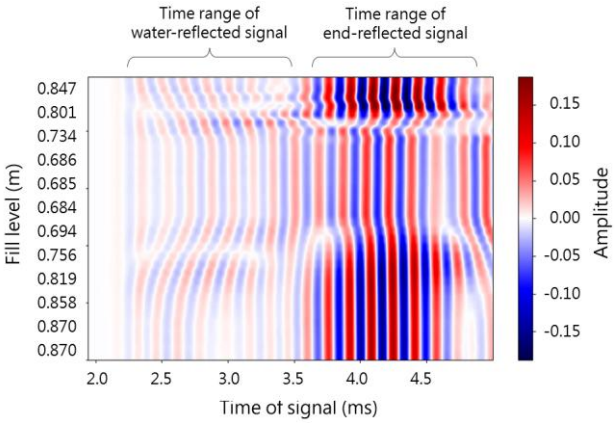


Figure 3: Time domain heatmap of measured guided wave signals showing the time ranges oft the water reflection and the end reflection for varying liquid fill levels.

Both the water reflected signal and the end reflected signal exhibit systematic temporal shifts as a function of the liquid fill level. When the fill level decreases from 0.87 m to 0.68 m, the water reflected signal shows a delay of up to 0.25 ms. Since this reflection propagates only along the air coupled section of the waveguide, the phase velocity remains constant and the observed time shift results solely from the increased propagation distance in air.
In contrast, the end reflected signal exhibits a smaller shift of approximately 0.04 ms towards earlier arrival times. Although the total propagation distance remains constant, the relative lengths of the air coupled and water coupled sections vary with fill level, leading to a change in the effective average propagation velocity and thus to a fill level dependent arrival time.

By jointly evaluating both reflections while accounting for the medium dependent propagation velocities, the liquid fill level can be determined relative to the hydrostatic pressure reference. The resulting deviation from the reference measurement remains below 2.0 cm (Figure 4).

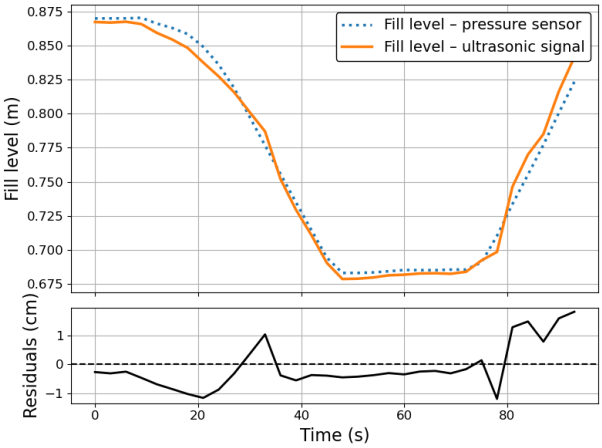


Figure 4: Comparison of water fill level determined from guided wave measurements and a hydrostatic pressure sensor, including the corresponding residuals.

3.2 Evaluation of internal phase boundaries

Building on the validated fill level evaluation, the capability of guided acoustic waves to resolve internal phase boundaries within stratified liquids is investigated in the following. In contrast to homogeneous fluid loading, layered fluid systems introduce additional acoustic impedance contrasts along the waveguide, which result in multiple distinct reflections in the guided wave signal. The simulations contain different time domain signals for an oil-water layered system at different liquid fill levels, corresponding to the conditions investigated in chapter 3.1 (Figure 5).

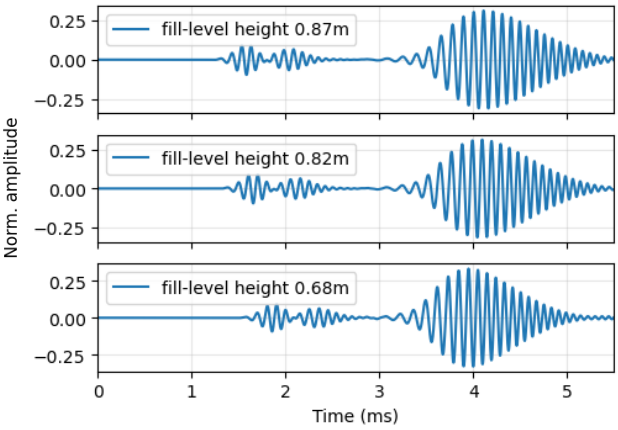


Figure 5: Simulated guided wave signals for an oil-water layered system at different total fill levels and an oil-layer thickness of 0.2 m.

In addition to the reflection from the upper oil surface, which corresponds to the global liquid fill level, a second reflection is observed that originates from the interface between the oil and water layers. The temporal separation between these two reflections is governed by the thickness of the oil layer and the effective propagation velocity of the guided wave within the oil loaded section of the waveguide. The final reflection corresponds to the end reflection after propagation along the complete waveguide length.
By evaluating the time of flight difference between the reflection from the oil surface and the reflection from the oil–water interface, the thickness of the internal oil layer can

be determined independently of the total fill level. In the present simulations, the oil layer thickness was defined as 0.2 m and serves as a reference value for assessing the accuracy of the proposed evaluation approach.
The resulting decomposition of the total liquid fill level into water, oil, and air contributions is shown in Figure 6.

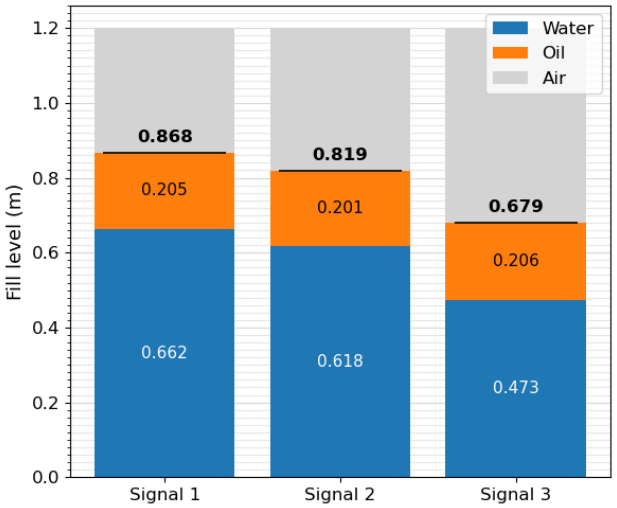


Figure 6: Determined liquid fill level and internal oil-layer thickness.

The oil layer thicknesses reconstructed from the guided wave signals show good agreement with the predefined simulation values for all investigated fill levels. At the same time, the water layer height adapts consistently with the variation of the overall fill level, confirming the internal consistency of the evaluation method.
In contrast to approaches based solely on attenuation or amplitude variations, the presented method relies on geometrically interpretable propagation times. As a result, the evaluation is inherently robust against moderate changes in excitation amplitude and damping behavior. This characteristic makes the approach particularly suitable for applications involving heterogeneous or multiphase fluids, where conventional level measurement techniques often fail to resolve internal interfaces reliably.

4 Conclusion

In this work, a polymer based ultrasonic guided wave sensor concept for liquid level monitoring in heterogeneous and stratified fluids was investigated. The presented results demonstrate that longitudinal and flexural guided wave modes provide reliable and consistent information on the global liquid fill level, while numerical simulations and experimental measurements show very good agreement, confirming the validity of the employed modeling approach.
By evaluating multiple reflections in the time domain, both the total liquid height and the thickness of internal layers can be reconstructed in a geometrically interpretable manner, resulting in a measurement approach that is inherently robust against amplitude variations and damping effects.
In contrast to conventional metallic waveguide sensors, the use of a high density polyethylene waveguide provides high chemical resistance and enables reliable operation in chemically aggressive media while maintaining a mechanically simple and cost efficient sensor design. The presented single sensor configuration further reduces system complexity compared to multi sensor measurement solutions, since fill level and stratification information can be obtained using the same compact sensing element.
Future work will focus on the experimental excitation and evaluation of the torsion mode for quantitative temperature compensation and on the validation of a multimodal sensor concept.

5 Literatur

[1] K. Lutter, A. Backer, and K. S. Drese. “Guided Acoustic Waves in Polymer Rods with Varying Immersion Depth in Liquid”. In: Sensors 23 (2023), p. 9892. DOI: 10.3390/s23249892.

[2] N. Subhash and B. Krishnan. “Fluid level sensing using ultrasonic waveguides”. In: Insight 56 (2014), pp. 607–612. DOI: 10.1784/insi.2014. 56.11.607.

[3] J. Shah, S. El-Hawwat, and H. Wang. “Guided Wave Ultrasonic Testing for Crack Detection in Polyethylene Pipes: Laboratory Experiments and Numerical Modeling”. In: Sensors 23 (2023), p. 5131. DOI: 10.3390/s23115131.

[4] J. S. Egerton et al. “Ultrasonic attenuation and phase velocity of high-density polyethylene pipe material”. In: The Journal of the Acoustical Society of America 141 (2017), pp. 1535–1545. DOI: 10.1121/1.4976689.

[5] M. Piechaczek, A. Backer, T. Roß, S. Tietze, and K. S. Drese. “Development of a continuous ultrasonic sensor for monitoring the fill level of existing separating layers”. In: Proceedings of the 9th International Congress on Ultrasonics (ICU), Paderborn, Germany (2025). DOI: 10.5162/Ultrasonic2025/A12-b4.

[6] A. Backer, V. Fairuschin, and K. S. Drese. “On Dispersion Compensation for GAW-Based Structural Health Monitoring”. In: Sensors 23 (2023), 4282. DOI: 10.3390/s23094282.

[7] J. Landskron, A. Backer, C. R. Wolf, G. Fischerauer, and K. S. Drese. “FEM-Based Modeling of Guided Acoustic Waves on Free and Fluid-Loaded Plates”. In: Appl. Sci. 15 (2025), 9116. DOI: 10.3390/app15169116.

DOI 10.5162/iCCC2026/P4

Contributions of Shifted Excitation Raman Difference Spectroscopy to Agri-Photonics – Pathway towards Smart Farming

Kay Sowoidnich, André Müller, Martin Maiwald
Ferdinand-Braun-Institut (FBH), Berlin, Germany, martin.maiwald@fbh-berlin.de

Abstract

Shifted excitation Raman difference spectroscopy (SERDS) provides molecule-specific information and can effectively address interferences from fluorescence and daylight. We applied SERDS using an in-house developed 785 nm dual-wavelength diode laser as excitation light source for soil investigations. Results from laboratory-based measurements as well as on-site field trials conducted with an in-house developed portable SERDS system demonstrate the capability for spatially-resolved qualitative and quantitative analysis of relevant soil parameters, including organic carbon and carbonates, thus providing important contribution to precision agriculture.

1 Introduction

The concept of precision agriculture [1] gains significant attention in modern agricultural practice. To derive informed treatment decisions for site-specific farming, e.g., with respect to fertilization, liming or irrigation, detailed information regarding the current soil composition is of crucial importance. Existing complex, expensive, and time-consuming standard laboratory analyses could ideally be complemented by fast and non-destructive optical methods that show great promise as innovative approach for soil characterization. Its ability to obtain molecule-specific information from soil makes Raman spectroscopy a well-suited candidate for such a task. However, due to the weakness of the Raman effect, the technique suffers from dominant background interferences, that can originate, e.g., from fluorescence or ambient daylight.

In this contribution, we apply shifted excitation Raman difference spectroscopy (SERDS) [2] as an effective tool to address such challenges. The technique uses a physical approach with two slightly shifted laser wavelengths that are used to record two Raman spectra. While Raman signals directly follow the applied wavelength shift, background interferences remain unaffected. Subtraction of the two recorded spectra thus provides a neat way of separating Raman signals from unwanted disturbing interferences.

We present initial SERDS investigation on collected soil samples using a customized laboratory setup. Subsequently, a portable SERDS instrument was developed and realized. This system was then applied for on-site soil investigations directly on an agricultural field in Northeast Germany to evaluate its capability for qualitative and quantitative analysis.

2 Materials and Methods

For all presented investigations, an in-house developed monolithic dual-wavelength diode laser with two emission lines around 785 nm [3] is applied as excitation light source for SERDS.

2.1 SERDS Laboratory setup

In preparation for on-site soil analysis, initial SERDS laboratory investigations were performed using 150 soil samples that were collected from an agricultural field in Germany. To address soil heterogeneity at the micrometer scale, a measurement spot size of 100 µm was chosen. The heterogeneity at the millimeter scale was addressed by applying a 100-point raster scan approach where each sample was probed in a 10×10-point grid within an area of approximately 1 cm^2. Further information can be found in our previous publications [4, 5].

2.2 Portable SERDS instrument

The transfer of SERDS from the laboratory environment towards the field setting is achieved by means of an in-house realized rugged and portable SERDS system [6] that was specifically designed for soil analysis. Similar to the above-mentioned laboratory setup and for comparison, also the portable instrument applies a measurement spot size of approximately 100 µm. For the on-site study on an agricultural field in Germany, twenty-five positions on the soil surface down to depths of 10 cm along a length of 480 m were exemplarily selected and at each position 10 different points were probed to address soil heterogeneity [7]. Additionally, soil samples were collected from the measurement positions for subsequent analysis with the SERDS laboratory setup to validate the results obtained on-site.

3 Results and Discussion

3.1 SERDS laboratory measurements

Results demonstrate the detection of organic carbon and 13 soil minerals, including silicates, carbonates, titanium dioxides, and phosphate. Moreover, a successful prediction of the contents of organic matter [4] and carbonate [5] as important soil parameters could be realized by means of partial least squares regression (PLSR). Overall, the outcomes of the laboratory study show the ability of SERDS

for qualitative (detection of soil constituents) and quantitative (prediction of organic matter and carbonate contents) soil investigations. In the next step, our approach was transferred from the laboratory towards the field environment by applying an in-house developed portable SERDS instrument.

3.2 On-site soil investigations

On-site investigations on the above-mentioned field were exemplarily carried out at selected points along a distance of 480 m. As an example, Raman spectra of one measurement position recorded at the two slightly shifted excitation wavelengths are displayed in Figure 1. It becomes evident that the characteristic Raman signals are superimposed on a strong background contribution. In this case, only the strong Raman signals of quartz (465 cm^{-1}) and the sapphire window contained in the Raman probe head (418 cm^{-1}) are hardly visible. Applying SERDS efficiently separates the Raman signals from the interfering backgrounds thus enabling the calcium carbonate polymorph aragonite (1084 cm^{-1}) and organic carbon (1350 cm^{-1} and 1590 cm^{-1}) to be detected and identified. Furthermore, two additional Raman signals originating from the sapphire window are present at 577 cm^{-1} and 749 cm^{-1}.

In terms of qualitative soil analysis, the detection of organic carbon and in total 9 soil minerals, e.g., silicates, carbonates, titanium dioxides, and phosphate was demonstrated thus confirming the results of the investigations using the SERDS laboratory setup. The measurements conducted at 25 selected points along a distance of 480 m revealed a spatially heterogeneous distribution of carbonaceous matter and carbonates across the field. As depicted in Figure 2, elevated concentrations of both constituents are located in a region covering distances of 250-350 m from the first probed point on the field.

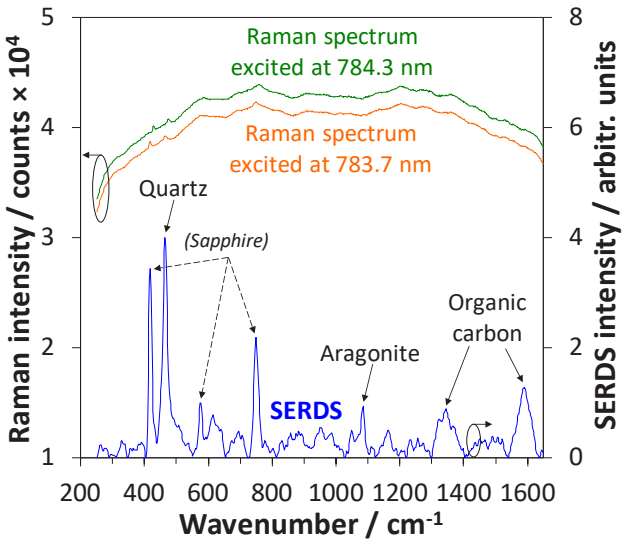


Fig. 1 Average of 50 Raman spectra excited at 783.7 nm and 784.3 nm using integration times of 50 ms for each Raman spectrum obtained on-site at a selected measurement position on an agricultural field in Germany (top curves), and reconstructed SERDS spectrum (bottom curve).

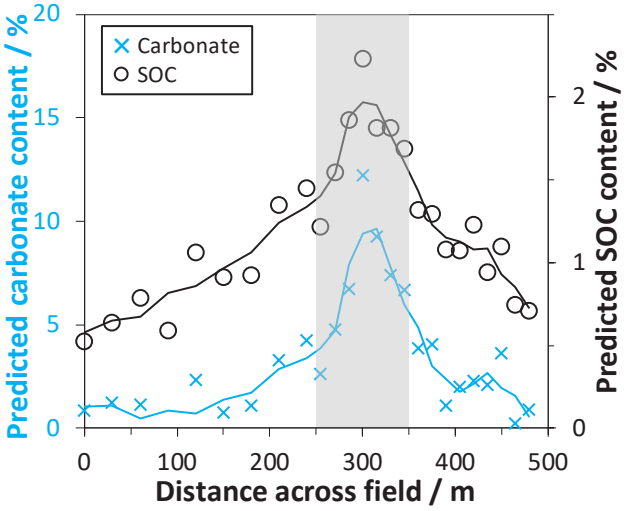


Fig. 2 Contents of carbonate and soil organic carbon (SOC) predicted by PLSR from SERDS spectra measured on-site with the portable SERDS instrument plotted in dependence of the measurement position across the field. Solid lines represent 3-point moving averages to guide the eye. The gray area indicates a range of elevated carbonate and SOC concentrations.

Moreover, a successful prediction of the contents of organic carbon (coefficient of determination: $R^2 = 0.89$, root mean square error of cross-validation: RMSECV = 0.3 %) and carbonate ($R^2 = 0.86$, RMSECV = 2.5 %) as important soil parameters could be realized. The achieved prediction performance is comparable to the SERDS laboratory analysis conducted on samples collected from the 25 probed locations on the field [7].

4 Conclusion

The outcomes of our study demonstrate the capability of portable SERDS for qualitative as well as quantitative on-site soil investigations directly on an agricultural field. This approach could significantly contribute to precision agriculture by providing spatially-resolved molecule-specific information about the soil composition.

5 Acknowledgement

This work was supported by the Federal Ministry of Research, Technology and Space (BMFTR) within the projects RaMBo (grant numbers 031A564C, 031B0513C and 031B1069C) within the consortium I4S (Intelligence for Soil) in the frame of the funding measure BonaRes (Soil as a Sustainable Resource for the Bioeconomy) and Research Fab Microelectronics Germany – FMD (grant number 16FMD02).

We would also like to thank all involved colleagues from I4S for sample collection and preparation, and the landowner Golo Philipp for access to the agricultural field where the on-site investigations were conducted.

6 References

[1] Bhakta, I.; Phadikar, S.; Majumder, K.: State-of-the-art technologies in precision agriculture: a systematic review. J. Sci. Food Agric. (2019) 99, 4878-4888.

[2] Shreve, A.P.; Cherepy, N.J.; Mathies, R.A.: Effective Rejection of Fluorescence Interference in Raman Spectroscopy Using a Shifted Excitation Difference Technique. Appl. Spectrosc. (1992) 46, 707-711.

[3] Müller, A.; Maiwald, M.; Sumpf, B.: Electrically wavelength adjustable DBR laser for background-free Raman spectroscopy at 785 nm. Proc. SPIE (2022) 12021, 1202103.

[4] Sowoidnich, K.; Vogel, S.; Maiwald, M.; Sumpf, B.: Determination of Soil Constituents Using Shifted Excitation Raman Difference Spectroscopy. Appl. Spectrosc. (2022) 76, 712-722.

[5] Sowoidnich, K.; Maiwald, M.; Ostermann, M.; Sumpf, B.: Shifted excitation Raman difference spectroscopy for soil component identification and soil carbonate determination in the presence of strong fluorescence interference. J. Raman Spectrosc. (2023) 54, 1327-1340.

[6] Maiwald, M.; Sowoidnich, K.; Sumpf, B.: Portable shifted excitation Raman difference spectroscopy for on-site soil analysis. J. Raman Spectrosc. (2022) 53, 1560-1570.

[7] Sowoidnich, K.; Pätzold, S.; Ostermann, M.; Sumpf, B.; Maiwald, M.: Qualitative and quantitative soil characterization on an agricultural field using a portable shifted excitation Raman difference spectroscopy instrument. Analyst (2025) 150, 2934-2944.

DOI 10.5162/iCCC2026/P5

Electrical Characterization of CeO_2/SnO_2 Hydrogen Sensors: Influence of Temperature on Baseline Sensor Currents

Amrutha Mudundi[1], Amanpreet Kalra[2], Carlos Morales[3], Jan Ingo Flege[3], Inga Anita Fischer[2], Carlos Alvarado Chavarin[1,*], Christian Wenger[1]
[1] IHP – Leibniz-Institut für innovative Mikroelektronik, Im Technologiepark 25, 15236 Frankfurt (Oder), Germany
[2] BTU Cottbus-Senftenberg, Angewandte Physik und Halbleiterspektroskopie, Erich-Weinert-Straße, 03046 Cottbus, Germany
[3] BTU Cottbus-Senftenberg, Experimentalphysik und funktionale Materialien, Erich-Weinert-Straße, 03046 Cottbus, Germany
*alvarado@ihp-microelectronics.com

Abstract

Growing environmental and energy sustainability challenges are driving increased interest in hydrogen as an emerging clean energy carrier and practical alternative to fossil fuels. Effective monitoring of H_2 concentrations in air is a critical requirement to ensure the safe operation of H_2-based technologies. Metal oxide (MO_x)-based resistive sensing is widely investigated, but their elevated working temperatures, often exceeding 300°C, lead to substantial increases in power consumption and reduce device longevity. Therefore, achieving H_2 sensing at or near room temperature is desirable for durable and scalable use. In this work, we report the use of CeO_2 and SnO_2-based heterostructures as potential sensing materials capable of lowering the operating temperature for H_2 detection in air. With this objective, we characterise the electrical behavior of the fabricated sensor devices under varying voltage and temperature conditions to assess baseline sensor currents. The gas sensing performance is also evaluated upon exposure to a 1% H_2/air mixture.

1 Introduction

In recent decades, significant efforts have been made to establish a sustainable energy economy to address global climate change and the depletion of fossil fuel reserves. In this regard, hydrogen, owing to its abundance and high energy efficiency, has emerged as a reliable and accessible green energy fuel. However, its wide explosion limits in air (4%-75%) under ambient conditions pose safety challenges that hinder large-scale commercialization [1]. As a consequence, reliable monitoring of hydrogen concentrations in air is crucial, driving the development of novel sensing materials and sensor prototypes. Numerous gas sensing mechanisms are based on physicochemical interactions between the target gas and the sensor's active layer [2]. Among all, resistive sensors based on reactive metal oxides are widely investigated due to their ease of fabrication, compatibility with integrated circuits, and high sensitivity. However, their operating temperatures are typically above 300°C, which increases power consumption and affects their long-term use [3]. Therefore, our current work focuses on reducing the working temperatures required to detect hydrogen in air.

The main working principle of resistive gas sensing relies on changes in the electrical current of the MO_x active layer upon exposure to analyte gas molecules. Based on this mechanism, several oxides and their combinations have been investigated and reported in the literature [4]. Nanostructured oxides, typically fabricated by techniques such as sol-gel, offer improved sensitivity due to their high surface-to-volume ratio. However, their deposition methods generally show poor compatibility with standard microelectronics processes, limiting their integration. As an alternative, in this work, we present thin-film deposition techniques for cerium oxide, tin oxide, and their heterostructures as post-processes compatible with CMOS fabrication. The metal-oxide channels are patterned and contacted using photolithography. The resulting sensor structures are electrically characterized, and the influence of temperature on the sensor's baseline currents is investigated. In the end, as a proof of concept, the current variation to 1% H_2/air mixture is also reported.

2 Device Fabrication

MO_x-based H_2 sensors typically have a simple structure, consisting of a sensing layer that interacts with the target gas and electrodes that act as transducing elements. Additionally, microheater platforms are incorporated into the sensor design to achieve elevated operating temperatures. In microfabrication, Atomic Layer Deposition (ALD) and Radio Frequency Magnetron Sputtering (RF-MS) are commonly employed to deposit MO_x thin films. MS is advantageous for producing relatively thick layers within short processing times, while ALD is preferred for precise thickness control and highly conformal coatings, with a slower growth rate. Earlier studies have demonstrated the deposition of CeO_2 active films using both MS [5] and ALD. Notably, ALD-grown CeO_2 has been shown to exhibit a measurable response to H_2 even at room temperature [6].

In our case, microsensor chiplets were fabricated, mounted on a customised microheater **(Figure 1a)**, and then wire-bonded to a Dual-In-Line package (DIL-24), as shown in **Figure 1d and 1e**. For the fabrication of CeO_2/SnO_2-based H_2 sensors, Si/SiO_2 substrates patterned with planarized interdigitated tungsten from IHP's cleanroom were used for

DOI 10.5162/iCCC2026/P5

deposition using a shadow mask. **(Figure 1b)**. SnO_2-doped CeO_2 layers of roughly 100 nm were deposited using a water-cooled 13.56 MHz RF-MS system over 60 minutes. A 99.99% (4N) CeO_2 target of 2" diameter with Sn wires on its surface was sputtered at an RF power of 10W. Ar (10 sccm) was used as the sputter gas at a working pressure of 10^{-2} mbar. Each fabricated sensor chiplet has two interdigitated electrode sets configured in a T-structure sharing a single common electrode. Further, for comparison, SnO_2-only channels were fabricated at IHP's Rapid Prototyping Lab (RPL) using maskless photolithography (Smart Print UV, Microlight 3D) as shown in **Figure 1c**. The process involved patterning the MO_x channel layer on SiO_2/Si substrates, followed by thermal evaporation of approximately 100 nm of Sn. The samples were then annealed in air at 300°C for an hour to oxidize the Sn. Finally, subsequent lithography steps were followed to pattern Ti/Au interdigitated electrode contacts.

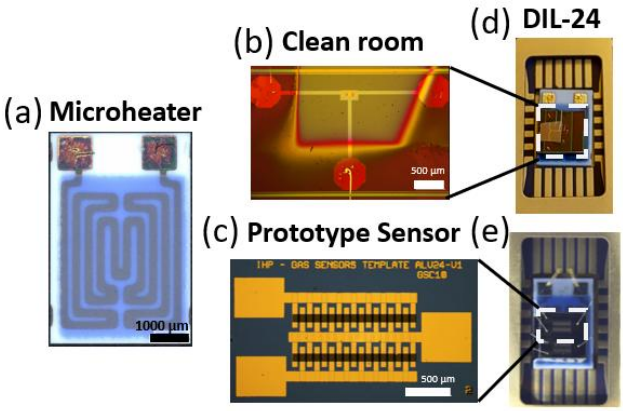


Figure 1 (a) Microheater, **(b)** CeO_2/SnO_2-based sensor chiplet, **(c)** SnO_2-based sensor chiplet, **(d and e)** sensor chiplets integrated with microheater and wire-bonded to a DIL-24 socket.

3 Electrical Characterization

CeO_2 exhibits versatile oxidation-state behavior under both reductive and oxidizing conditions, making it a highly suitable material for sensing applications. It is also being extensively investigated for the detection of H_2 at moderate operating temperatures, especially in heterostructures with more conductive oxides such as SnO_2 [7]. MO_x-based sensors exhibit time-dependent current variations due to aging and oxidation effects. Also, the electrical characterization of previous samples had a reduction in current by orders of magnitude. Current drifts in MO_x-based sensors have also been reported in literature [8] [9] , attributed to oxygen vacancies and intrinsic material properties. In the present work, electrical characterization of our sensors has been carried out to evaluate the baseline sensor behavior over time under the influence of temperature and H_2 exposure. This baseline characterization helps eliminate possible misinterpretations in sensor responses. The results have been compared between CeO_2/SnO_2 devices and SnO_2-only devices.

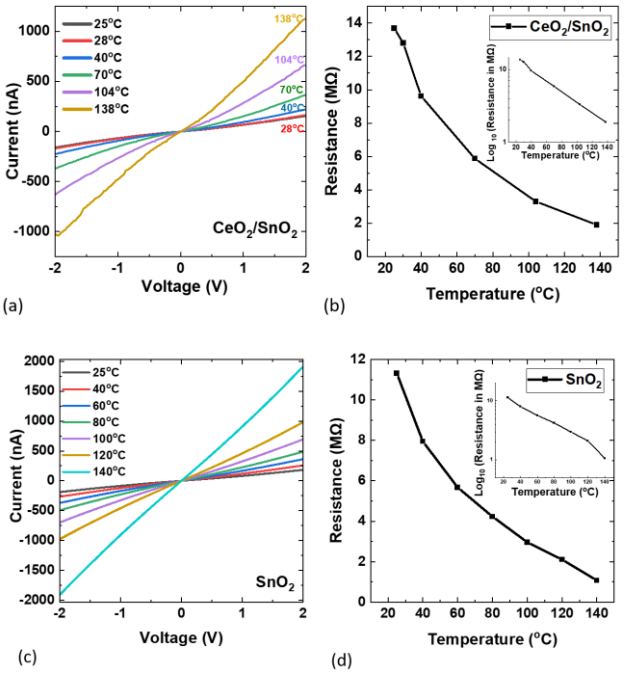


Figure 2 Effects of temperature on both CeO_2/SnO_2 and SnO_2-only structures. I-V characteristics with heating **(a)** CeO_2/SnO_2 (micro-heater voltage) and **(c)** SnO_2. **(b and d)** Resistance variation with temperature increase in CeO_2/SnO_2 and SnO_2-only structures, respectively. Inset: log (resistance) vs. temperature.

The influence of temperature on sensor current was evaluated based on current vs voltage (I-V) measurements. **Figure 2a** presents the I-V characteristics of the CeO_2/SnO_2 channels measured at increasing temperatures, controlled by varying the microheater voltage. Within the measured voltage range (-2 V to 2 V), all the I-V curves exhibit nearly linear behavior with a coefficient of determination (R^2) of approximately 0.99, indicating predominantly ohmic conduction. As the temperature increases, the additional thermal energy enhances charge transport, resulting in steeper I-V slopes. Consequently, the device resistance decreases significantly with temperature **(Figure 2b)**, with the value at 138°C (1.9 MΩ) dropping to nearly 15% of the room-temperature resistance (13.7 MΩ). A logarithmic plot of resistance versus temperature **(Figure 2b inset)** was inserted to clearly illustrate this thermally activated conduction. Similar measurements were carried out for the SnO_2-only channels using the PMV200 electrical characterization tool. To evaluate the effect of temperature, the samples were heated to the desired temperatures using the PMV200 heater. The resulting I-V curves **(Figure 2c)** showed a consistent trend with temperature, with currents increasing from 190 nA to 1900 nA across the measured range (25°C-140°C). **Figure 2d** confirms the material's semiconducting behavior via the non-linear decrease in resistance with temperature. Furthermore, the log (resistance) **(Figure 2d inset)** shows a sharp drop in resistance for only a 20°C rise in temperature between 120°C and 140°C, emphasizing the strong temperature dependence.

DOI 10.5162/iCCC2026/P5

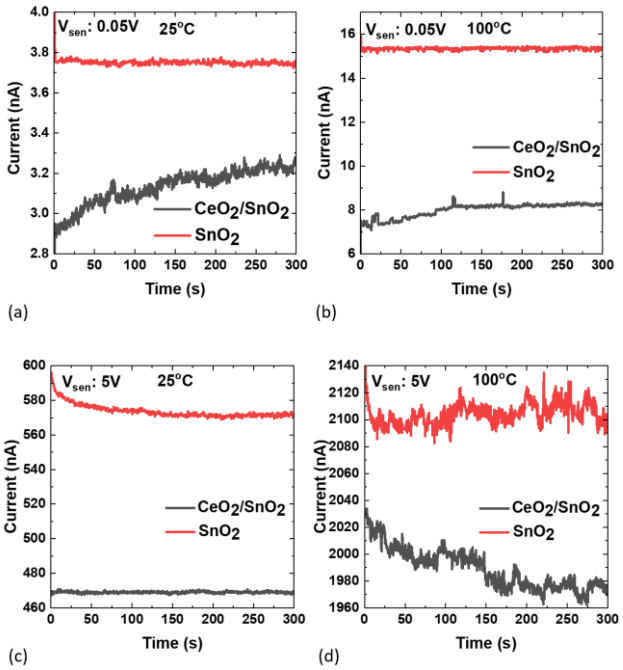


Figure 3 Observations of current drift in the CeO_2/SnO_2 channel (black) and the SnO_2-only channel (red) at low voltage (a and b) and high voltage (c and d) at room temperature and 100°C, respectively.

Condition	Current Drift (pA/s)	
	CeO_2/SnO_2	SnO_2-only
0.05 V, 25°C	+ 0.99	- 0.05
0.05 V, 100°C	+ 2.48	+ 0.61
5 V, 25°C	+ 2.03	- 81.92
5 V, 100°C	- 198.79	- 106.53

Table 1 Current drift values at each voltage and temperature condition for the CeO_2/SnO_2 and SnO_2-only channels.

A current drift analysis was carried out for both MO_x-based channels to assess current variations at very low (0.05 V) and relatively high (5 V) sensor voltages under different temperature conditions (25°C and 100°C) **(Figure 3)**. For this purpose, current values were recorded for 5 minutes to examine whether the currents stabilized over time. Based on the measurements, current drift values were calculated in each case and tabulated **(Table 1).** The CeO_2/SnO_2 channel shows a modest increase in current under low voltage operation, rising from around 1 pA/s at 25°C to higher values at elevated temperatures **(Figure 3a and 3b)**. At high voltage and 100°C, however, the drift becomes significantly negative (~ 199 pA/s), indicating reduced stability. In comparison, the SnO_2-only channel remains relatively stable at low voltages but shows pronounced negative drift at high voltage conditions**. (Figure 3c and 3d)**. Overall in both cases, while the current behavior under mild conditions remains fairly stable, there is substantial instability under high voltage and elevated temperature operation.

In the end, the current variation in the MO_x channels due to a 1% H_2/air gas stream was measured at temperatures suitable for the surrounding electronics **(Figure 4)**. As shown in **Figure 4a**, there is a 12% current variation upon gas exposure at 50°C for the CeO_2/SnO_2 channel, whereas the SnO_2-only channel **(Figure 4b)** exhibits a slightly lower change of 6% at 70°C. Although the current ratio between before-exposure (I_o) and after-exposure (I) to the H_2 mixture is relatively low compared to previously reported values, achieving a measurable response at these low working temperatures represents a step toward reduced power consumption. The literature reports an increase in current in the MO_x layer when exposed to a reducing gas such as H_2; however, our results show an unexpected decrease in measured current. The mechanism behind this current behavior remains unclear and is under ongoing investigation. In future work, ultra-thin (< 20 nm) MO_x films and their heterostructures will be deposited by ALD, and their performance will be compared with that of sputtered films.

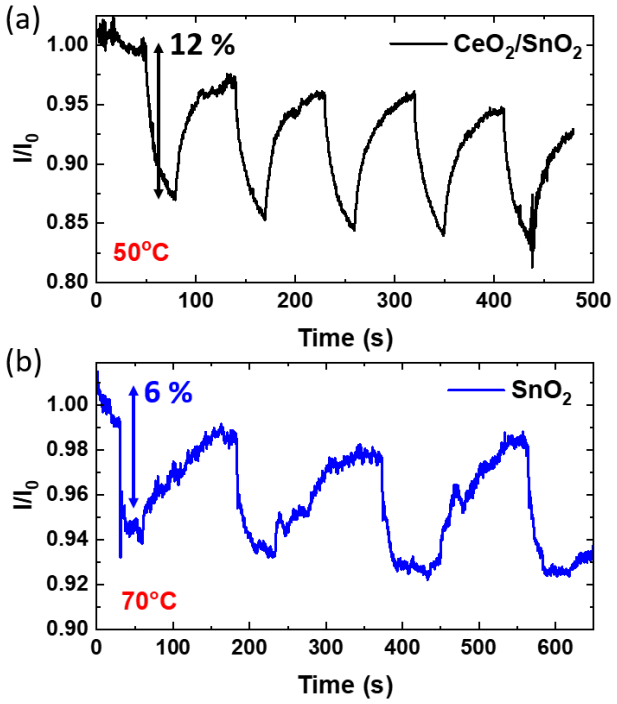


Figure 4 Current variation upon exposure to 1 % H_2/air gas mixture at different temperatures in **(a)** CeO_2/SnO_2 sample and **(b)** SnO_2-only sample.

4 Conclusion

In this study, we report a thin film deposition method for CeO_2/SnO_2 heterostructures as a post-processing step that can be integrated with CMOS procedures. For comparison purposes, SnO_2-based sensor device fabrication was also done following rapid prototyping procedures via maskless photolithography. The sensor chiplets were then integrated with a custom microheater and wire bonded to a DIL-24 socket for electrical characterization at elevated temperatures. I-V measurements reveal a significant temperature dependence with reduced resistance at high temperatures. The current drift analysis indicates notable instability under high voltage and high temperature operation. Finally, from the results of gas exposure to a 1% H_2/air mixture, with detectable current variation at 50°C, the CeO_2/SnO_2-based device demonstrates better performance than the

SnO_2-only device. Although current changes remain low, these baseline results provide a reference for the electrical behavior and guide further optimization of the MO_x active layer and sensor performance.

5 Acknowledgements

This work was funded by Germany's Federal Ministry of Research, Technology and Space (BMFTR) under grant No. 16ME0424.

6 Literature

[1] T. Hübert, L. Boon-Brett, W. Buttner, *Sensors for Safety and Process Control in Hydrogen Technologies*. in 2018. CRC Press.

[2] T. Hübert, L. Boon-Brett, G. Black, and U. Banach, "Hydrogen sensors – A review," *Sens. Actuators B Chem.*, vol. 157, no. 2, pp. 329–352, Oct. 2011, doi: 10.1016/j.snb.2011.04.070.

[3] P. Wang, S. Xu, X. Shi, J. Zhu, H. Xiong, and H. Wen, "Recent Advances in Resistive Gas Sensors: Fundamentals, Material and Device Design, and Intelligent Applications," *Chemosensors*, vol. 13, no. 7, p. 224, Jun. 2025, doi: 10.3390/chemosensors13070224.

[4] N. Goel, K. Kunal, A. Kushwaha, and M. Kumar, "Metal oxide semiconductors for gas sensing," *Eng. Rep.*, vol. 5, no. 6, p. e12604, Jun. 2023, doi: 10.1002/eng2.12604.

[5] A. Hayat *et al.*, "Structural and morphological properties of CeO2 films deposited by radio frequency magnetron sputtering for back-end-of-line integration," *Thin Solid Films*, vol. 807, p. 140547, Oct. 2024, doi: 10.1016/j.tsf.2024.140547.

[6] C. Morales *et al.*, "Hydrogen Sensing via Heterolytic H_2 Activation at Room Temperature by Atomic Layer Deposited Ceria," *ChemSusChem*, vol. 18, no. 13, p. e202402342, Jul. 2025, doi: 10.1002/cssc.202402342.

[7] D. E. Motaung *et al.*, "Ultra-high sensitive and selective H2 gas sensor manifested by interface of n–n heterostructure of CeO2-SnO2 nanoparticles," *Sens. Actuators B Chem.*, vol. 254, pp. 984–995, Jan. 2018, doi: 10.1016/j.snb.2017.07.093.

[8] G. Müller and G. Sberveglieri, "Origin of Baseline Drift in Metal Oxide Gas Sensors: Effects of Bulk Equilibration," *Chemosensors*, vol. 10, no. 5, p. 171, May 2022, doi: 10.3390/chemosensors10050171.

[9] N. Dennler, S. Rastogi, J. Fonollosa, A. van Schaik, and M. Schmuker, "Drift in a popular metal oxide sensor dataset reveals limitations for gas classification benchmarks," *Sens. Actuators B Chem.*, vol. 361, p. 131668, Jun. 2022, doi: 10.1016/j.snb.2022.131668.

DOI 10.5162/iCCC2026/P6

Nanostrukturierte Dünnschicht-Schottky-Fotodioden für SWIR-Fotodetektoren

Leonor Shala[1,2)], Jens Knobbe[1)], Lion Augel[1,2)],
[1)]Fraunhofer IPMS (Institutsteil ISS), Cottbus, DE
[2)]Lehrstuhl für Mikro- und Nanosysteme, Brandenburgische Technische Universität, Cottbus, DE
leonor.shala@ipms.fraunhofer.de

Kurzfassung

Silizium-basierte Schottky Fotodioden liefern einen Schnittpunkt zwischen Fotodetektion im SWIR-Bereich und einer CMOS-Kompatibilität, welche sie zu einem Bauelement von aktuellem Interesse macht. Mittels Verwendung einer Dünnschicht im Bereich von 10 – 20 nm als Schottky-Metall sowie einer pyramidalen Nanostrukturierung des Substrates können in vielversprechender Weise die interne sowie externe Quanteneffizienz verbessert werden. Ob man das Sputtern oder die pCVD als Abscheidemethode zum Aufbringen des Schottky-Metalls auf das Substrat wählt, hat einen Einfluss auf die Qualität der Schottky-Barriere. Der Einfluss der Abscheidemethode, der Schichtdicke des Schottky-Metalls sowie der Strukturierung des Substrats auf die Schottky-Barrierequalität wird mittels temperaturabhängiger Strom-Spannungs-Messungen analysiert.

Abstract

Silicon based Schottky Photodiodes offer detection in the SWIR range as well as a CMOS compatibility. Through introduction of a thin film in the range of 10 – 20 nm as Schottky-metal, as well as a pyramidal nanostructuring of the substrate, one can improve the quantum efficiency of the device in a promising manner. The choice of either sputtering or pCVD as deposition method for the Schottky-metal has a direct influence on the quality of the Schottky-barrier. In this work, the change of barrier quality through differing film thickness, deposition method and substrate structuring is examined through temperature dependent I-V measurements.

1 Motivation

In den vergangenen Jahrzehnten hat sich der „komplementäre Metall-Oxid-Halbleiter“ (engl. CMOS) Prozess als Standard-Technologie im Bauelement-Design etabliert. Dies geht so weit, dass in der Halbleiter-Industrie die CMOS-Kompatibilität von neuartigen Technologien von entscheidender Bedeutung über den Fortschritt bzw. die weitere Verwendung einer solchen Technologie ist. Dazu kommt getreu „Moore’s Law“ die stetige Miniaturisierung der Bauelement-Strukturen, wodurch die Verwendung unter anderem von Dünnschichten im Bereich weniger Nanometer favorisiert wird. Die in diesem Bericht behandelten Dünnschicht-Schottky-Fotodioden folgen diesen zwei Schlüsselkompetenzen. Zum einen gelingt die CMOS-Kompatibilität dadurch, dass direkt auf Silizium als Substrat strukturiert und beschichtet wird. Zum anderen wird auf das Si-Substrat ein Dünnschicht-Metall als Schottky-Kontakt abgeschieden.

1.1 Si- und Schottky-basierte SWIR-Sensorik

Neben der CMOS-Kompatibilität ist Silizium als Substratmaterial ebenfalls aufgrund der Transparenz zu Licht mit Wellenlängen < 1,2 µm vom Vorteil. Denn die „Kurz-Wellenlängen Infrarot“ (engl. SWIR) Sensorik befasst sich mit dem Wellenlängenbereich 1,4 – 3 µm. Somit kann ein rückseitenbeleuchtetes System in Betracht gezogen werden, in welchem das Licht durch das Si-Substrat hindurch zum aktiven Bereich (hier dem Schottky-Kontakt) gelangt. Der Schottky-Kontakt wird durch das Abscheiden von Titan-Nitrid (TiN), welches aufgrund seiner vergleichsweise niedrigen Zustandsdichte nahe dem Ferminiveau favorisiert wird [1], als Dünnschicht auf der Si-Oberfläche geformt. Durch das Substrat hindurch gelangt das Licht zur Metallschicht, wo es mit seiner Energie einen Ladungsträger anregt. Dieser angeregte Ladungsträger (engl. „hot carrier“) kann nun über den Schottky-Übergang vom Metall in den Halbleiter emittieren, sofern seine Energie mindestens die der Barrierenhöhe ist, welche sich zwischen dem Metall und dem Halbleiter gebildet hat. Emittierte Ladungsträger können final als elektrisches Signal abgegriffen und verarbeitet werden.

1.2 Dünnschicht-Metallisierung

Die Güte eines Fotodetektors wird zu einem wesentlichen Teil durch seine interne Quanteneffizienz (IQE) bestimmt. Hierbei handelt es sich im Fall eines Schottky-Detektors um das Verhältnis der Anzahl erfolgreich über den Schottky-Kontakt emittierter Ladungsträger zu der Anzahl an Photonen einer bestimmten Energie, welche vom Detektorsystem (hier dem Schottky-Metall) absorbiert wurden und einen „hot carrier“ erzeugt haben. Findet demnach jeder durch eingehende Photonen erzeugte „hot carrier“ mittels Schottky-Emission den Weg in den Halbleiter, liegt die IQE folglich bei 1.

DOI 10.5162/iCCC2026/P6

Ein „hot carrier“ kann neben der Schottky-Emission auch thermisch oder mittels Elektron-Phonon-Streuung emittieren. Die hierbei freigewordene Energie trägt nicht zur Schottky-Interaktion bei und verringert folglich die IQE. Das statistische Auftreten dieser Effekte wird u.a. durch die mittlere freie Weglänge eines Ladungsträgers beschrieben. Ist das Verhältnis der mittleren freien Weglänge zur Schichtdicke des Metalls aufgrund einer hohen Schichtdicke gering, so ist die Schottky-Emissionswahrscheinlichkeit ebenfalls gering, da der „hot carrier“ vor einer Schottky-Emission wahrscheinlich anderweitig (thermisch, phononisch) emittiert. Um die Schottky-Emissionswahrscheinlichkeit (und damit auch die IQE) zu erhöhen, kann man daher die Schichtdicke des Metalls reduzieren. [2]

1.3 Pyramidale Nanostrukturierung

Neben der IQE ist in Anbetracht des Gesamt-Bauelements auch die externe Quanteneffizienz (EQE) relevant. Diese beschreibt das Anzahlverhältnis Schottky-emittierter Ladungsträger zu allen eingehenden Photonen. Werden keine eingehenden Photonen durch das Bauelement reflektiert oder transmittiert, sondern alle Photonen absorbiert („hot carrier“-Erzeugung), so entspricht die EQE der IQE. Eine verringerte Reflektion und Transmission aus dem Bauelement heraus erhöht somit die EQE.
Um Reflektionen an der Schottky-Zwischenfläche aus dem System heraus zu minimieren, werden in das Si-Substrat vor der Metallabscheidung sog. „Pyramidenstrukturen“ geätzt, wie in Bild 1 zu sehen ist. Diese Strukturierung sorgt im Vergleich zu einer planaren Substratoberfläche dafür, dass eingehende Photonen zwar weiterhin an der Zwischenfläche reflektiert werden, diese reflektierten Photonen aber durch die „Einbuchtung“ weiterhin mittels Mehrfachreflektion die Möglichkeit haben, absorbiert zu werden. Zudem begünstigt die nun nanostrukturierte Metallschicht bei eingehenden Photonen Effekte wie die „Oberflächenplasmonenresonanz“ (engl. SPR), welche die verstärkte Absorption von Photonen bestimmter Frequenz durch im Nanometerbereich beschränkte freie Elektronen mit passender Schwingungsfrequenz beschreibt. [3]
Semi-transparente Oxidschichten auf der Rückseite des Substrats sowie auf dem Schottky-Metall minimieren zusätzlich Reflektionen und Transmissionen aus dem System heraus.

2 Methodik

Die Herstellung der Fotodioden erfolgt in Vielzahl auf 200-mm-Wafern im Reinraum des Fraunhofer IPMS in Dresden, welcher auf industriellem Fabrikationsstandard operiert. Die nachgehende Charakterisierung der Fotodioden erfolgt im Wesentlichen am Institutsteil ISS des Fraunhofer IPMS, welches in Cottbus ansässig ist.

2.1 Herstellung

Die wesentlichen Vorgänge in der Fabrikation sind die Ätzung der Pyramidenstrukturen sowie die Abscheidung des Schottky-Dünnschichtmetalls.
Bei der Strukturätzung werden mittels reaktivem Ionenätzen (engl. RIE) in zwei Stufen isotrop Wölbungen in das Si-Substrat geätzt, woraus sich dann eine Pyramidenform ergibt. Diese haben einen Abstand (Spitze zu Spitze) von 400 – 800 nm und eine Breite (Pyramidenboden) von 1000 – 1500 nm.

2.2 Abscheidung

Die Abscheidung des Schottky-Metalls erfolgt entweder durch Sputtern (Niederschlag des zuvor zerstäubten Metalls auf Substrat) oder durch eine gepulste chemische Gasphasenabscheidung (engl. pCVD; Abwechselnde „Eintrittspulse“ eines Präkursors und eines Ko-Reaktanten sowie Reaktion zum gewünschten Schichtmaterial auf dem Substrat). Während das Sputtern grundsätzlich mit weniger Prozessaufwand verknüpft ist und auf planaren Substraten hochqualitative Schichten erzeugt, hat pCVD den Vorteil, dass es auch auf strukturierten Substraten homogene und konforme Schichten erzeugt [4].
Im Sputterprozess werden sowohl TiN-Schichten als auch TiN/AlSiTi-Schichten als Schottky-Metallschicht abgeschieden. Im pCVD-Prozess wird lediglich TiN abgeschieden.

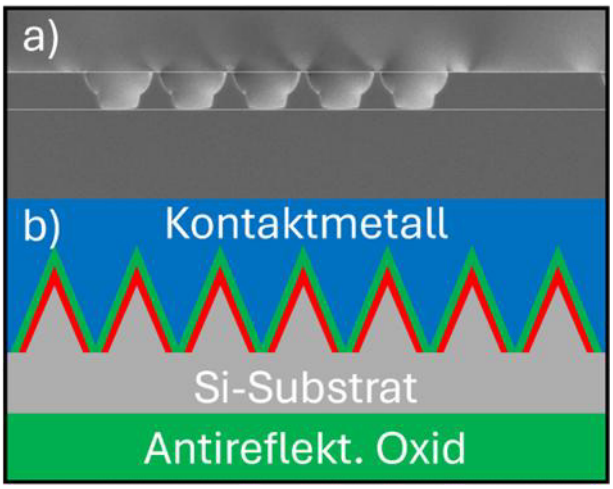


Bild 1 a: Querschnittsaufnahme der pyramidalen Strukturen mittels REM, b: Schema des Schichtaufbaus der Fotodiode. Auf dem Pyramidal strukturierten Substrat wird vor der Metallisierung (blau) das Schottky-Metall (rot) sowie eine Oxidschicht (grün) abgeschieden.

2.3 Charakterisierung

Durch temperaturabhängige Strom-Spannungs-Messungen kann Aufschluss über die Barriere-Eigenschaften des Schottky-Übergangs erlangt werden (Ladungsträgertransport, Barriere-Inhomogenitäten).
Es werden im Folgenden die Unterschiede im Strom-Spannungs-Kurvenverlauf bei verschiedenen Temperaturen (T=233/253/273/293/313 K) zwischen Proben mit unterschiedlichen Abscheidemethoden des Schottky-Metalls

DOI 10.5162/iCCC2026/P6

(Sputtern, pCVD), unterschiedlichen Schottky-Metallschichtdicken (10 nm/20 nm) und unterschiedlicher Strukturierung (planar/pyramidal) verglichen.

3 Ergebnisse

3.1 Abscheidemethode

Im direkten Vergleich der Abscheidemethoden (Sputtern siehe Bild 2a; pCVD siehe Bild 2b) bei gleicher Dicke von 10 nm und planarer Struktur stellt man zunächst fest, dass der Strom im Falle der pCVD-Probe um etwa eine Größenordnung geringer ist als der der Sputter-Probe. Bei Schichtdicken von 10 nm ist es vor allem im Sputterprozess möglich, dass über die Diodenfläche lokale Störstellen am Kontakt der Substratoberfläche mit dem Schottky-Metall (Versetzungen, Korngrenzen) auftreten, welche Möglichkeiten für Leckströme bieten. Räumlich betrachtet ist die Kontaktfläche zwischen Substrat und Schottky-Metall nicht flach, sondern rau, wodurch es grundsätzliche räumliche Fluktuationen in der Barrierenhöhe gibt. Dies erklärt den grundsätzlichen linearen Anstieg des Stroms beider Proben bei niedrigeren Spannungen [5]. Der leichte, nicht-lineare Anstieg des Stroms der Sputter-Probe bei T=233 Kelvin suggeriert, dass Tunnelströme vom Schottky-Metall in den Halbleiter existieren.

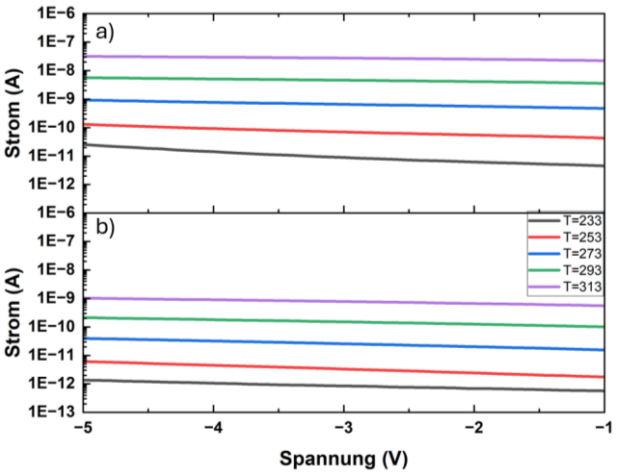


Bild 2 Vergleich der Sperrströme bei Temperaturen von 233 – 313 Kelvin (siehe Inlet). a: Planare Probe mit 10 nm Schottky-Metallschicht, abgeschieden durch Sputtern. b: Planare Probe mit 10 nm Schottky-Metall, abgeschieden durch pCVD.

3.2 Schichtdicke

Dies wird bei Betrachtung von Bild 3 deutlich, in welchem bei gleicher Abscheidung (Sputtern) und planarer Struktur die Unterschiede in den Schichtdicken (10 nm siehe Bild 2a; 20 nm siehe Bild 2b) verglichen werden. Vor allem der nicht-lineare Anstieg bei T=233 Kelvin, aber auch der grundsätzliche lineare Anstieg aller Kurven wird durch die Erhöhung der Schichtdicke minimiert. Dies lässt sich durch eine längere Beschichtungsdauer und somit einer gleichmäßigeren Beschichtung begründen, welche Störstellen wie z.B. Freistellen am Schottky-Kontakt oder auch in der Schottky-Metallschicht minimiert.

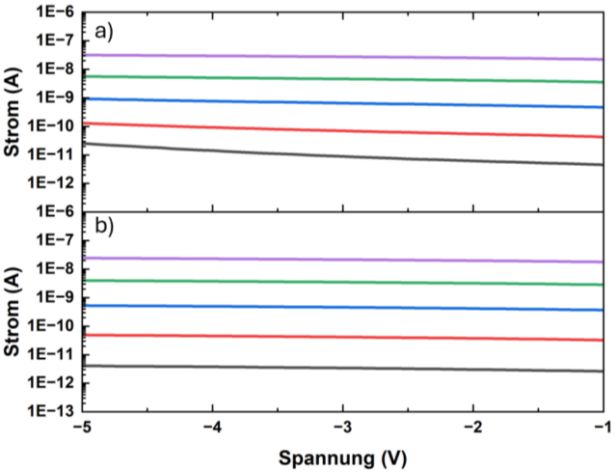


Bild 3 Vergleich der Sperrströme bei Temperaturen von 233 – 313 Kelvin (siehe Inlet Bild 2). a: Planare Probe mit 10 nm Schottky-Metallschicht, abgeschieden durch Sputtern. b: Planare Probe mit 20 nm Schottky-Metallschicht, abgeschieden durch Sputtern.

3.3 Strukturierung

Bei gleicher Schichtdicke (20 nm) und Abscheidung (Sputtern) werden die Proben bei planarer Struktur (Bild 4a) und pyramidaler Strukturierung (Bild 4b) verglichen. Durch die Minimierung des Kurvenanstiegs über die Schichtdicke ist der Kurvenverlauf in diesem Vergleich nahezu identisch. Die im Vergleich zu pCVD inkonformere und inhomogenere Schicht bei strukturiertem Substrat fällt bei einer Schichtdicke von 20 nm allen Anschein nach nicht mehr wesentlich ins Gewicht.

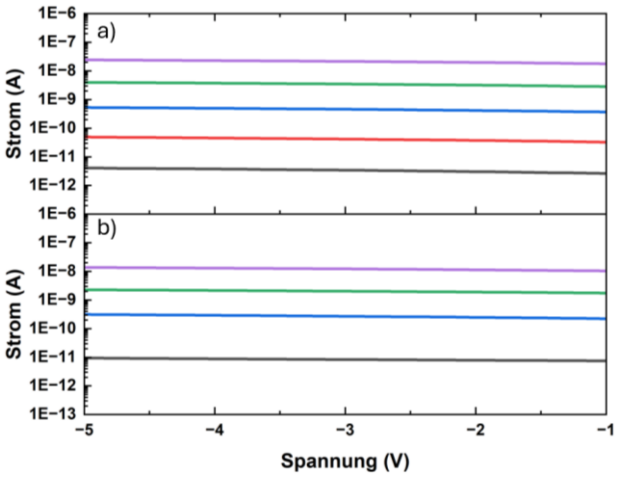


Bild 4 Vergleich der Sperrströme bei Temperaturen von 233 – 313 Kelvin (siehe Inlet Bild 2). a: Planare Probe mit 20 nm Schottky-Metallschicht, abgeschieden durch Sputtern. b: Pyramidal strukturierte Probe mit 20 nm Schottky-Metallschicht, abgeschieden durch Sputtern.

4 Zusammenfassung

In diesem Beitrag werden über temperaturabhängige Strom-Spannungs-Messungen Fotodioden mit unterschiedlichen Schottky-Metallschichtdicken, Schottky-Metallabscheidemethoden und Substratstrukturierungen verglichen, um einen Einfluss dieser Parameter auf die Schottky-Barrierenqualität zu ermitteln. Neben der Erhöhung der Schichtdicke hat auch die Wahl der pCVD als Abscheidemethode einen positiven Einfluss auf den Kurvenverlauf, da lineare sowie nicht-lineare Anstiege des Sperrstroms in negative Richtung minimiert werden. Ein Einfluss der Strukturierung des Substrats konnte bei erhöhter Schichtdicke nicht ermittelt werden.

Es gilt in Folgeversuchen nun, vor allem bei geringeren Schichtdicken die Unterschiede bei sich ändernder Strukturierung von Proben mit gesputterter sowie mit pCVD vollführter Schottky-Metallabscheidung zu untersuchen. Grundsätzlich wird jedoch nach den präsentierten Ergebnissen aus Bild 2 die pCVD als Abscheidemethode für zukünftige Versuche favorisiert.

5 Literatur

[1] AJ. Khurgin, „Fundamental limits of hot carrier injection from metal in nanoplasmonics”, Nanophotonics, vol. 9, no. 2, pp. 453-471, 2020

[2] J. Duran and A. Sarangan, “Schottky-barrier photodiode internal quantum efficiency dependence on nickel silicide film thicknesses”, IEEE Photonics Journal, vol. 11, no. 1, pp. 3-5, 2019

[3] H. Wen; L. Augel; J. Knobbe „Approaches for Optimizing Near Infrared Si Photodetectors Based on Internal Photoemission”, ICICDT, 2021

[4] D.J.H. Emslie; P. Chadha, J.S.Price „Metal ALD and pulsed CVD: Fundamental reactions and links with solution chemistry”, Coordination Chemistry Reviews, vol. 257, no. 23-24, pp. 3282-3296, 2013

[5] J.H. Werner and H.H. Güttler “Barrier inhomogeneities at Schottky contacts”, J. Appl. Phys. 69, 1522-1533, 1991

DOI 10.5162/iCCC2026/P7

Ammonium and Potassium Ion-Sensitive Field-Effect Transistors

L. T. Kühne[1], C. Ruffert[2,3], C. Beale[1], M. Wambold[1], E. Kurth[1], O. R. Hild[1]
[1] Fraunhofer Institute for Photonic Microsystems (IPMS), Maria-Reiche-Str. 2, 01109 Dresden, Germany
[2] Fraunhofer Institute for Photonic Microsystems (IPMS), Konrad-Zuse-Str. 1, 03046 Cottbus, Germany
[3] Brandenburgische-Technische Universität Cottbus-Senftenberg
Corresponding author's e-mail address: olaf.hild@ipms.fraunhofer.de

Abstract

The integration of potassium-selective membranes on ion-sensitive field-effect transistors (ISFETs) was successfully reproduced in this study, confirming results previously reported [1]. A sensitivity of 60.2 mV/Dec was achieved in KH_2PO_4/NaH_2PO_4 buffer solution at pH 7, which is in close agreement with the value of 59.3 mV/Dec reported in earlier work. This outcome demonstrates the capacity for stable and reproducible membrane integration, with the potential to approach the theoretical Nernstian response. Furthermore, the coefficient of determination R^2 for all sensors was consistently above 99.75 %, indicating excellent linearity and reliability of the sensor signal. However, measurements performed in a KCl solution buffered with tris(hydroxymethyl)aminomethane to pH 7.6 revealed a significantly reduced sensitivity of 49.36 mV/Dec, despite similar experimental conditions and a still high R^2-value above 99.5 %. This discrepancy suggests a noticeable influence of interfering sodium ions on the sensor's performance, particularly regarding the selectivity of the potassium-sensitive membrane. A comparison of the NH_4^+ sensor reveals a lower sensitivity. The sensitivity is 25.11 mV/Dec and exhibits both a slightly increased variance and a reduced coefficient of determination.

1 Introduction

Ion-sensitive field-effect transistors (ISFETs) are quite promising microdevices for the continuous monitoring of ion activity in biomedical processes. The device's key advantages are its small size, low impedance, and low cost when produced on a large scale. Potassium (K^+) is known to activate numerous enzymatic processes. Consequently, even minor fluctuations in the concentration of K^+ may result in severe ailments, such as arrhythmia and other cardiovascular complications [2-4]. Potassium (K^+) is a vital element for all living cells. In plants it regulates various processes, including growth processes, protein synthesis and stomata regulation, which facilitates gas exchange [5]. Potassium is also of relevance in the field of agriculture, where it plays a crucial role in the synthesis of plant protein and cellulose. In addition, it serves to regulate water balance and control root turgor [6, 7]. The importance of accurate, direct, fast, and cost-efficient K^+-concentration detection capabilities is hence obvious.

Baccar et al. investigated K^+ ISFET microsensors fabricated by ion implantation. The K^+ sensitivity, selectivity, drift, reproducibility, and the lifetime were investigated. The sensitivity of the implanted ISFET equal to 41-44 mV/Dec was still stable after a year's tests with a response range of 10^{-3} to 10^{-1} M [8]. The drift was determined to be smaller than 0.01 mV/h. Two types of sensors were used: **I**) electrolyte-insulator-semiconductor (EIS) structures and **II**) ISFET microsensors using valinomycin as ionophore. The K^+ ISFETs featured an interference for high concentrations of sodium (Na^+) ions.

Benslimane et al. focused on three primary objectives: **I**) to analyse potassium ion levels in Moroccan soils utilizing an ISFET with a selective polyvinyl chloride membrane, **II**) to evaluate the efficacy of different soil-to-water ratios as extraction methods, and **III**) to propose a novel and straightforward extraction technique [9]. The results indicated that the microsensor ISFET provided accurate results on solutions prepared in the laboratory with a correlation factor of 90 % and R^2 greater than 80 % when estimating potassium levels in soil extracts. The potassium concentration in an aqueous solution was determined by a polyvinyl chloride (PVC) membrane based on valinomycin, tetrakis(4-chlorophenyl)borate as an ionophore, bis(2-ethylhexyl) sebacate (DOS) as a plasticizer, and cyclohexanone as a solvent, using an integrated Ag/AgCl reference electrode. The response range was 5×10^{-4} to 10^{-1} M, and the sensitivity was 48 mV/Dec.

A graphene-based ISFET for potassium sensing was proposed and demonstrated by Fakih et al. [10]. Real-time sensing of K^+ was achieved with a detection limit of 10^{-9} M, and a resolution of $\sim2\times10^{-3}$ log[K^+]. The ISFETs exhibited long-term stability with limited drift over five months. The cross-sensitivity resulted in 2.5 mV/Dec for Na^+, 4.2 mV/Dec for Ca^{2+}, 1.5 mV/Dec for Mg^{2+}, and 9.0 mV/Dec for NH_4^+. Experiments with a variety of specimens confirmed the suitability of graphene ISFETs for K^+ sensing in fluids with multiple solutes.

Riedel et al. [11] found that PVC-membrane ISFETs for K^+, NH_4^+, NO_3^-, and $H_2PO_4^-$ (K^+ and NH_4^+ ≈ 50 mV/Dec; NO_3^- ≈ 48 mV/Dec; $H_2PO_4^-$ ≈ 42 m/Dec) have only a slight long-term drift manifested predominantly as a parallel shift of calibration curves (offset ~0.03 - 0.04 V over three weeks) with unchanged slope.

Hamlaoui et al. [12] reported a zeolite-modified ammonium ISFET using a siloprene membrane incorporating clinop-tilolite. The sensor yielded a sensitivity of ~32 mV/pNH_4^+ and a detection limit of 10^{-8} M. Interferences from H^+, Na^+, and K^+ were negligible (<3 mV/Dec). The device demonstrated stable operation without clinoptilolite leaching for over one month.

In this study, we present a feasibility test of K^+ and NH_4^+ ISFET microsensors as an extension to a previous publication [1].

DOI 10.5162/iCCC2026/P7

2 Materials and Method

2.1 Ionophore-ISFET

As a basis for ionophore sensors, we used Ta_2O_5-ISFETs from the 200 mm wafer production line at Fraunhofer IPMS. The p-channel ISFETs were produced on n-type (1-2 Ωcm) 200 mm (100) Si-wafers with dry SiO_2 gate oxide. For the pH-sensitive layer, about 150 nm Ta_2O_5 were deposited onto the gate oxide by r.f. sputtering.

Before coating the ISFETs, they were cleaned with ethanol (15 min) and distilled water. 3 µl of potassium- (with cyclopentanone) or ammonium- (with THF) ionophore cocktail, produced by Aptisens Ltd, were drop-casted and dried for an hour at a maximum temperature of 45 °C. In the drying process, the layer profile becomes inhomogeneous, which means that the layer thickness above the gate may vary. By performing profile measurements, we found that the PVC membrane usually has a crater-like shape, with steep edges and a flatter, more even middle area. Therefore, it was assumed that the diffusion potential is approximately constant over the ISFET channel area.

2.2 Electrical Characterisation

The fabricated ISFETs were operated in constant-voltage, constant-current (CVCC) mode for the measurements. Up to 10 ISFETs could be measured simultaneously in the setup shown in **Figure 1**. The sensors share a common solid-state Ag/AgCl reference electrode, which is mounted into the lid of the measurement chamber. However, for chloride-containing solutions, due to the chloride sensitivity of the solid-state reference electrode, an Ag/AgCl glass reference electrode was used.

The temperature within the measuring vessel was regulated by a thermostat jacket, which was utilized to maintain a constant temperature of 20 °C during the measurements. Using a mechanical stirrer ensured a uniform temperature throughout the electrolyte.

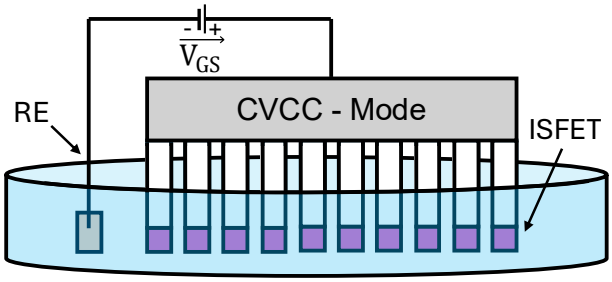


Figure 1 Measurement setup for up to 10 ISFET with a shared reference electrode system (RE).

For all measurements, the drain-source potential (V_{DS}) was set to -0.5 V, the drain-source current (I_{DS}) was set to -200 µA and the bulk-source potential (V_{BS}) was set to 0 V. To characterise the K^+ sensitivity we used **I)** a dilution series of a 1.5 M KCl solution buffered with 0.2 M tris(hydroxymethyl)aminomethane (Tris), adjusted to a pH of 7.6 with HCl and **II)** a dilution series of Merck Certipur® pH 7 buffer, which contains disodium phosphate and potassium dihydrogen phosphate [13]. For the NH_4^+ sensitivity, an ammonium nitrate dilution series was prepared, with each step lasting 30 min in every measurement.

The K^+ sensors were conditioned in a 1 mmol/l KNO_3 and NaCl solution, while the NH_4^+ sensors were conditioned in a 0.4 mmol/l NH_4NO_3 solution for a period of 15 hours (overnight) prior to the measurements.

3 Results and Discussion

3.1 Potassium ISFET

First results with Fraunhofer IPMS ISFETs with a K^+-sensitive PVC layer were presented in the first paper [1] with a measured sensitivity of 59.3 ± 0.8 mV/Dec at 20 °C. To confirm this sensitivity and verify a suspected cross-sensitivity, further measurements were conducted in different environments.

In **Figure 2** several ISFETs with a K^+-sensitive layer and one pH-ISFET without a membrane as a reference are presented. The measurement from 1.5 mol/l to 1.5×10^{-4} mmol/l KCl is pH-buffered with tris(hydroxymethyl)aminomethane (Tris) to pH 7.6. The simultaneously measured pH value indicates a deviation at the lowest dilution, due to the very weak buffering effect.

The response of the K^+-sensitive ISFETs in each step is quick, but shows a slight drift, especially in higher dilution. This is to be expected, as the use of a salt bridge with a diaphragm for the reference electrode results in slight contamination with 3 M KCl. However, with uniform contamination, the K^+ slope should only be slightly affected, as each solution has the same amount of additional potassium ions after 30 minutes. Only the actual K^+ concentration used to calculate the linear regression is not precise.

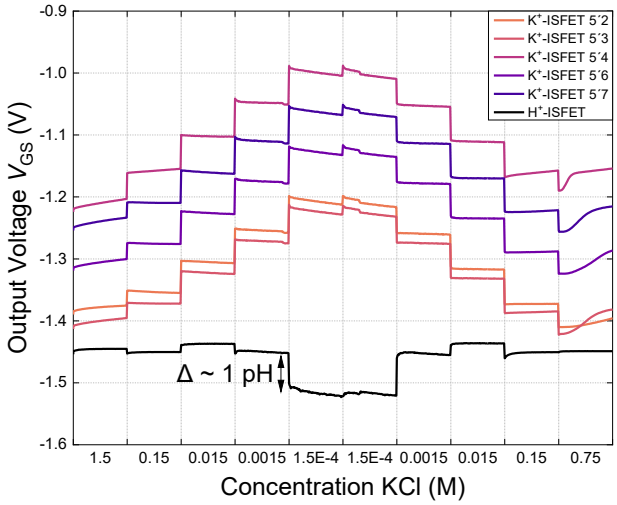


Figure 2 Potassium measurement from 1.5 to 1.5×10^{-4} mol/l KCl, pH-buffered with Tris and with an Ag/AgCl glass reference in 3 M KCl.

The analysis of the voltage values after a settling time of 30 minutes indicates a linear relationship between the sensor voltage and the negative decimal logarithm of the concentration in a range from 0.15 to 1.5×10^{-4} mol/l KCl,

DOI 10.5162/iCCC2026/P7

as given in **Figure 3**. The concentration range was cropped for evaluation, as the high concentrations fall outside the specifications of the ionophore cocktails.

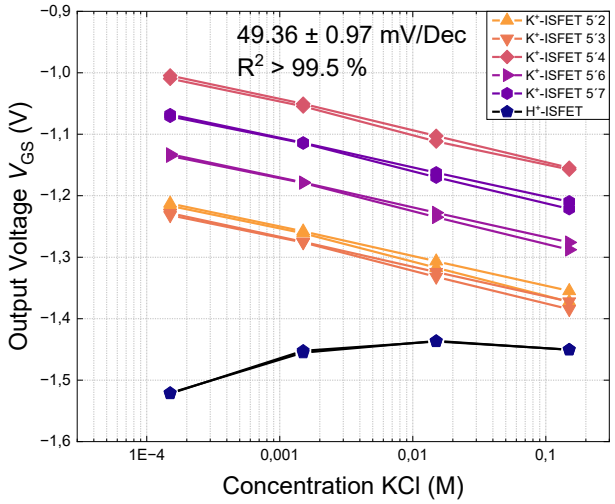


Figure 3 Linear sensitivity to KCl in a concentration range of 150 - 0.15 mmol/l for five representative K^+-selective ISFETs and one pH-ISFET.

The slope of all K^+-sensing ISFET is 49.36 ± 0.97 mV/Dec at 20 °C, in the concentration range from 150 to 0.15 mmol/l KCl. Each K^+ ISFET has a Coefficient of determination R^2 of over 99.5 %.

As already discussed in Sec. 2.1, that the layer thickness of the sensors is expected to exhibit slight inhomogeneity, this represents excellent reproducibility of K^+ sensitivity. The minor variations in sensitivity would be rectified through calibration in later applications, underscoring the significance of achieving a high level of linearity.

For the second potassium measurement, sensitivity was evaluated in a disodium phosphate and potassium dihydrogen phosphate solution. The Merck Certipur® pH 7 buffer was diluted from 1/1 to 1/1000 for this purpose. The fitted slope of 60.20 ± 1.06 mV/Dec for the three K^+-ISFETs (**Figure 4**) is slightly above the Nernstian slope at 20 °C.

These values reproduce prior measurements in KH_2PO_4/NaH_2PO_4 solution, which exhibited a sensitivity of 59.3 ± 0.8 mV/Dec [1] and are consistent with the theoretical Nernst slope for monovalent ions. This outcome demonstrates the capacity for stable and reproducible membrane integration with minimal variation between ISFET sensors apart from baseline offsets. Furthermore, the R^2-value for all sensors was consistently above 99.75 %, indicating excellent linearity and reliability of the sensor signal after calibration.

However, since the measurements performed in a KCl solution buffered with Tris to pH 7.6 revealed a significantly reduced sensitivity of only 49.36 mV/Dec, despite identical test conditions and a still high R^2-value of over 99.5 %, this discrepancy suggests a noticeable influence of interfering Na^+ ions on the performance of the sensor, particularly with regard to the selectivity of the potassium-sensitive membrane. These observed cross-sensitivities are consistent with the results of Riedel et al. [11], who specified a logarithmic selectivity coefficient ($K_{A,B}^{Pot}$) of -3.8 for Na^+ ions for K^+ ionophores.

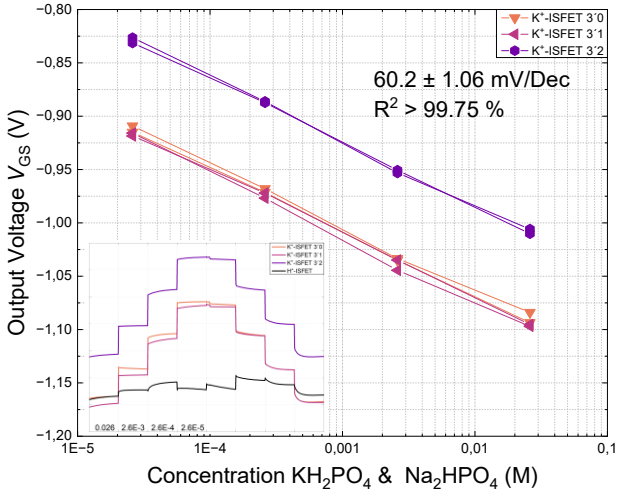


Figure 4 Sensitivity for three K^+-selective ISFETs in a KH_2PO_4/NaH_2PO_4 buffer dilution series at pH 7 measured against a solid-state Ag/AgCl reference.

In conclusion, although the membrane integration is both highly effective and reproducible, the ionic composition of the measurement medium is found to be a crucial factor in determining the sensitivity of the sensor. This highlights the necessity to take cross-sensitivities into account during future calibrations and in application-specific sensor designs.

3.2 Ammonium ISFET

In addition to the successfully integrated K^+ membranes, initial ionophore membranes with ammonium (NH_4^+) were investigated, with a focus on expanding knowledge in the general field of ionophore integration.

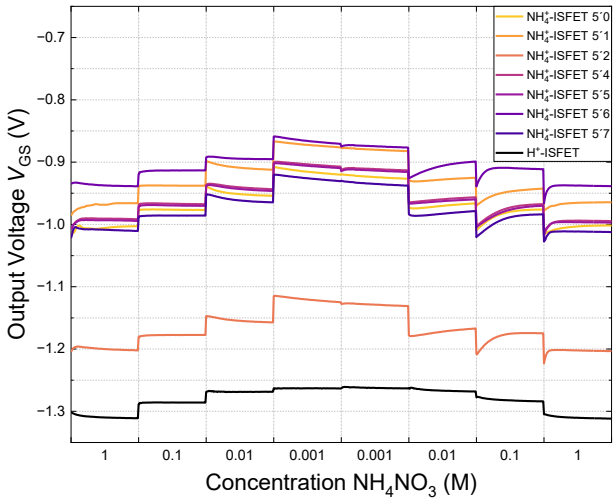


Figure 5 Ammonium measurement from 1 to 1×10^{-3} mol/l of NH_4NO_4, measured against a solid-state Ag/AgCl reference.

Figure 5 provides the measurement curves for seven representative NH_4^+ ISFETs. Additionally, one pH-ISFET

is given for reference purposes. In the dilution series, the ammonium nitrate concentration of the solution is first reduced from 1 to 1×10^{-3} mol/l, before gradually increasing back to the initial concentration.

The pH-ISFET demonstrates a slight change in pH value during dilution due to changes in ionic strength and the dissociation equilibrium. These fluctuations have been previously observed in unbuffered solutions and are thus not considered in further discussion. In comparison, the NH_4^+ ISFETs demonstrate a more pronounced increase in output voltage, with clear steps at each dilution stage. A sensor drift is observable in the individual steps, however in contrast to the K^+ ISFET against an Ag/AgCl reference electrode in 3 M KCl, this is not attributable to contamination with the measurement ions or cross ions.

Nevertheless, the slopes in **Figure 6** are 25.11 ± 3.44 mV/Dec, which is only half as steep as that observed in comparable K^+ ISFETs. The R^2-value is found to be acceptable for all NH_4^+ ISFETs at over 96.39 %.

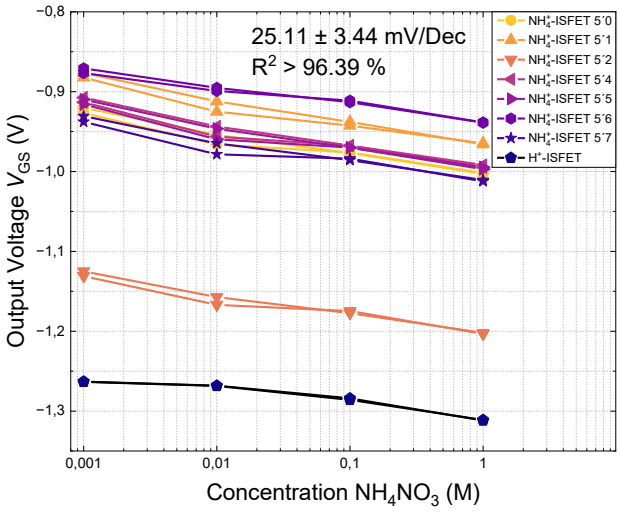


Figure 6 Sensitivity for NH_4^+ selective ISFETs in a NH_4NO_3 dilution series.

4 Conclusion

This work demonstrates the feasibility of integrating K^+ and NH_4^+ selective PVC membranes onto Ta_2O_5-ISFETs, and analyses sensitivity, linearity, and probable cross-sensitivity. Despite the expected inhomogeneity of the membrane thickness, the sensors exhibit strong reproducibility, indicating that the effective diffusion potential across the active gate region remains uniform enough to enable reliable sensing.

Multiple ISFETs with K^+ sensing layers are measured in KCl and KH_2PO_4/NaH_2PO_4 buffer solutions. The slope varied depending on the presence of certain cross ions. A reproducible slope of 59.3 and 60.2 mV/Dec, respectively, was observed in the presence of Na^+ in the dilution series. However, without Na^+ ions, only a slope of 49.36 mV/Dec was observed in a buffered KCl solution. Although different reference electrodes were used for certain reasons, the unmodified ISFETs show that the reference potential does not affect the measurement. This confirms that K^+ ionophores are sensitive to Na^+ ions.

A comparison of the two membranes shows that the NH_4^+ sensitive membranes exhibit a significantly lower sensitivity. The slope is 25.11 mV/Dec and exhibits both a slightly increased variance and a reduced coefficient of determination.

5 Literature

[1] Kühne, L. T., Wambold, M., Beale, C., Al-Falahi, F., Kurth, E., Hild, O. R. u. Ruffert, C.: Ion-Sensitive Field-Effect Transistors for Potassium Sensing. MikroSystemTechnik Kongress 2025. 2025

[2] He, F. J. u. MacGregor, G. A.: Beneficial effects of potassium on human health. Physiologia Plantarum 133 (2008) 4, S. 725–735

[3] Hughes Jr., F. M. u. Cidlowski, J. A.: Potassium is a critical regulator of apoptotic enzymes in vitro and in vivo. Advances in Enzyme Regulation 39 (1999) 1, S. 157–171

[4] Ravens, U. u. Cerbai, E.: Role of potassium currents in cardiac arrhythmias. EP Europace 10 (2008) 10, S. 1133–1137

[5] Leigh, R. A. u. Wyn Jones, R. G.: A hypothesis relating critical potassium concentrations for growth to the distribution and functions of this ion in the plant cell. New Phytologist 97 (1984) 1, S. 1–13

[6] Jackson, D. T. u. Nelson, P. N.: Preparation and properties of some ion selective membranes: A review. Journal of Molecular Structure 1182 (2019), S. 241–259

[7] de Kaiser, Lamb, J. A. u. Rosen, C.: Plant analysis sampling and interpretation University of Minnesota Extension FO-3176-B (2013)

[8] Baccar, Z. M., Jaffrezic-Renault, N., Martelet, C., Jaffrezic, H., Marest, G. u. Plantier, A.: K^+-ISFET type microsensors fabricated by ion implantation. Materials Chemistry and Physics 48 (1997) 1

[9] Benslimane, O., Rabie, R., Diane, A., Bensaleh, M., Saidi, O., Bouzida, I. u. El Hajjaji, S.: Evaluation of Soil Potassium Using Miniaturized Electrochemical Sensors. Ecological Engineering & Environmental Technology 25 (2024) 10, S. 206–215

[10] Fakih, I., Centeno, A., Zurutuza, A., Ghaddab, B., Siaj, M. u. Szkopek, T.: High resolution potassium sensing with large-area graphene field-effect transistors. Sensors and Actuators B: Chemical 291 (2019), S. 89–95

[11] Riedel, V., Hinck, S., Peiter, E. u. Ruckelshausen, A.: Handling of Ion-Selective Field-Effect Transistors (ISFETs) on Automatic Measurements in Agricultural Applications Under Real-Field Conditions. Electronics 13 (2024) 24

[12] Hamlaoui, M., Kherrat, R., Marrakchi, M., Jaffrezic-Renault, N. u. Walcarius, A.: Development of an ammonium ISFET sensor with a polymeric membrane including zeolite. Materials Science and Engineering: C 21 (2002) 1, S. 25–28

[13] Buffer Calibration Standards for pH Measurements MK_BR8716EN, Merck KGaA, 2022

DOI 10.5162/iCCC2026/P8

IR Soil Moisture – Contactless Detection of Forest Soil Dryness for Early Wildfire Risk Assessment

Ute Franke, 5micron GmbH, Berlin / Germany, ute.franke@5micron.de

Kurzfassung

In der Landwirtschaft, Forstwirtschaft sowie bei der Bewirtschaftung von Torfgebieten ist es entscheidend, den Feuchtigkeitsgehalt der Bodenoberfläche und der obersten Pflanzenschicht zu kennen. Diese wenige Zentimeter dünne Schicht spielt eine zentrale Rolle bei der Regulierung der Bewässerung, der Pflanzenhygiene, der Bestimmung optimaler Erntezeitpunkte sowie bei der frühzeitigen Erkennung von Brandgefahren.

Unter bestimmten Wetterbedingungen kann die Bodenoberfläche stark austrocknen, während tiefere Bodenschichten noch feucht sind. Klassische Messverfahren wie kapazitive Sensoren, Radar-, Ultraschall- oder Satellitenmessungen erfassen in der Regel den Wassergehalt tieferer Schichten und liefern keine zuverlässigen Daten zur Oberflächenfeuchte.

Das innovative Messsystem von 5micron schließt diese Lücke. Durch den Einsatz vernetzter Sensoren – darunter Infrarot-, Temperatur- und Helligkeitssensoren – sowie KI-gestützter Datenanalyse ermöglicht es eine präzise Echtzeit-Bestimmung der Bodentrockenheit und des Feuchtigkeitsgehalts von Wald-, Acker- und Vegetationsflächen. Gefährdete Regionen können so frühzeitig identifiziert und gezielte Präventivmaßnahmen eingeleitet werden.

Das System trägt zum Schutz von Wäldern, Artenvielfalt und landwirtschaftlichen Erträgen bei, reduziert wirtschaftliche Schäden und erhöht die Sicherheit von Gemeinden in Risikogebieten.

Abstract

In agriculture, forestry, and peatland management, monitoring the moisture content of the soil surface and the upper plant layer is critical for irrigation management, plant health, and early fire prevention. 5micron's innovative measurement system provides real-time information on surface soil moisture through modulated infrared reflection sensors (980 nm) combined with AI-supported data analysis. The system delivers robust, drift-free readings both day and night and operates under ultra-low-power principles. Proof-of-concept studies demonstrated a measurement range of 0–30 % w/w with stable signal quality. Wireless data transmission enables centralized analysis incorporating site-specific parameters such as soil type, sensor height, and environmental conditions. The system allows early identification of vulnerable areas, enabling preventive measures, while contributing to forest and biodiversity protection, reducing economic damage, and enhancing community safety

1 Innovation

Zunehmende Trockenperioden und Wetterereignisse wie starke Winde erhöhen das Risiko von Waldbränden erheblich. Die Entflammbarkeit des Bodens hängt maßgeblich vom Trockenheitszustand der obersten Zentimeter ab, da hier die meisten Primärbrandherde entstehen.

Bestehende Verfahren zur lokalen Bodenfeuchtemessung sind häufig invasiv, wartungsintensiv oder nur punktuell einsetzbar. Flächenhafte Messsysteme basieren meist auf Simulationen oder Satellitendaten (z. B. des Deutschen Wetterdienstes) und liefern nur begrenzte Informationen über den Zustand der obersten Bodenschicht. Zudem ermöglichen Satelliten keine kontinuierlichen Datenströme und erfassen daher keine kurzfristigen Schwankungen durch Föhnlagen oder starke Winde. Flächenbezogene Daten geben außerdem keinen Aufschluss über den Trockenheitszustand an besonders kritischen Standorten wie Picknick- oder Grillplätzen sowie entlang von Bahn- oder Stromtrassen.

Das von 5micron entwickelte Endprodukt ist ein berührungsloses Infrarot-Sensorsystem zur kontinuierlichen Erfassung der lokalen, oberflächennahen Waldbodentrockenheit. Es ermöglicht eine frühzeitige und präzise Waldbrandrisikoerkennung auch an sensiblen Einzelstandorten.

2 Technologie

Das System basiert auf einer modulierten Infrarot-Reflexionsmessung im Wellenlängenbereich von 980 nm und erfasst gezielt die Feuchte der obersten Bodenschicht (ca. 3–5 cm). Physikalische Grundlage ist die unterschiedliche

Absorptionseigenschaft von Wasser: Eine feuchte Oberfläche absorbiert Infrarotlicht stärker als eine trockene und reflektiert entsprechend weniger Strahlung.

Durch den Einsatz modulierten Infrarotlichts ist das System weitgehend unabhängig vom solaren Infrarotanteil. Störende Einflüsse durch Sonnenlicht werden gezielt unterdrückt, wodurch eine stabile Messung sowohl im Tages- als auch im Nachtbetrieb gewährleistet wird.

Die Sensoreinheit basiert auf modulierten IR-LEDs und Fotodioden und ermöglicht eine berührungslose Messung aus einer Distanz von bis zu 3 Metern. Im Rahmen eines Proof of Concept wurde ein Messbereich von 0–30 % w/w bei stabiler Signalqualität nachgewiesen. Die Ergebnisse zeigen eine robuste und driftfreie Messperformance.

Die erfassten Reflexionsdaten werden zusammen mit Umgebungshelligkeit und Temperatur per Funk an einen zentralen Server übertragen und dort ausgewertet. In die Berechnung fließen standortspezifische Parameter wie Bodentyp, Sensorhöhe und Umgebungsbedingungen ein.

Abbildung 1 den Schnitt durch einen typischen Boden
Abbildung 2 zeigt das Prinzipschema der Messmethode,.

Abbildung 1: Messbereich

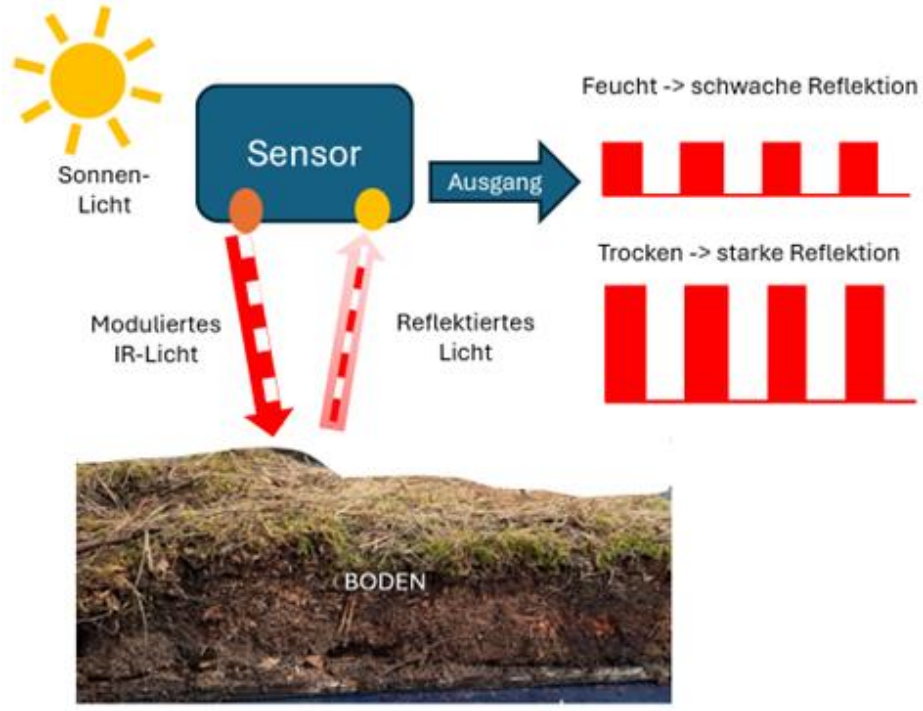
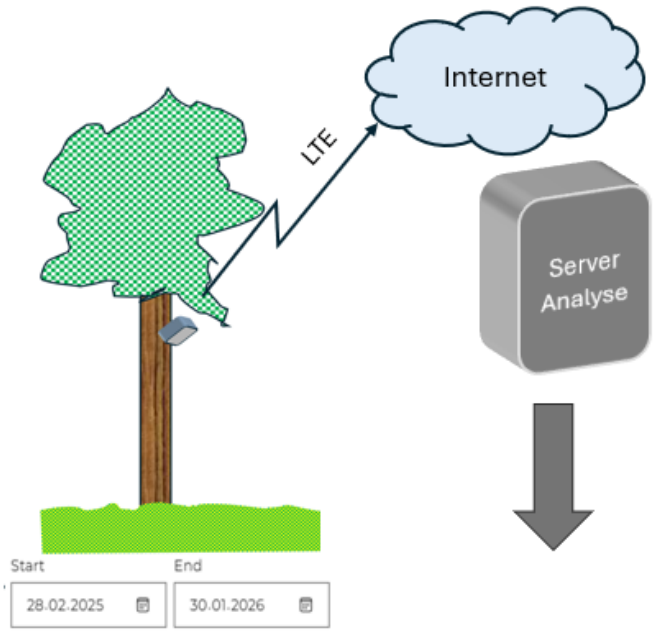
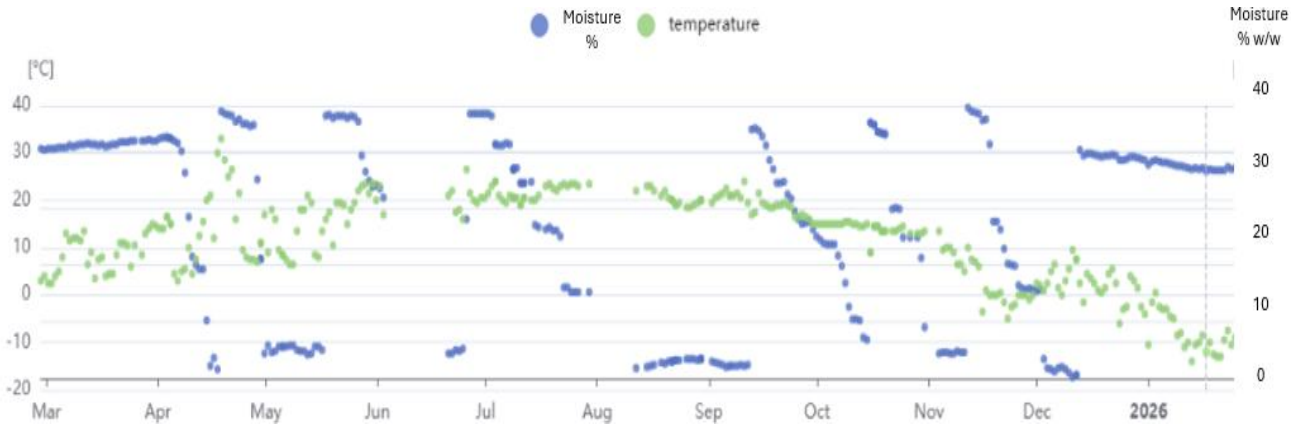


Abbildung 2: Prinzipschema der Messmethode

Abbildung 3 zeigt die Darstellung des kompletten Messkette und Beispielbild der Front-End Datenausgabe mit Zeit - Messreihe

DOI 10.5162/iCCC2026/P8





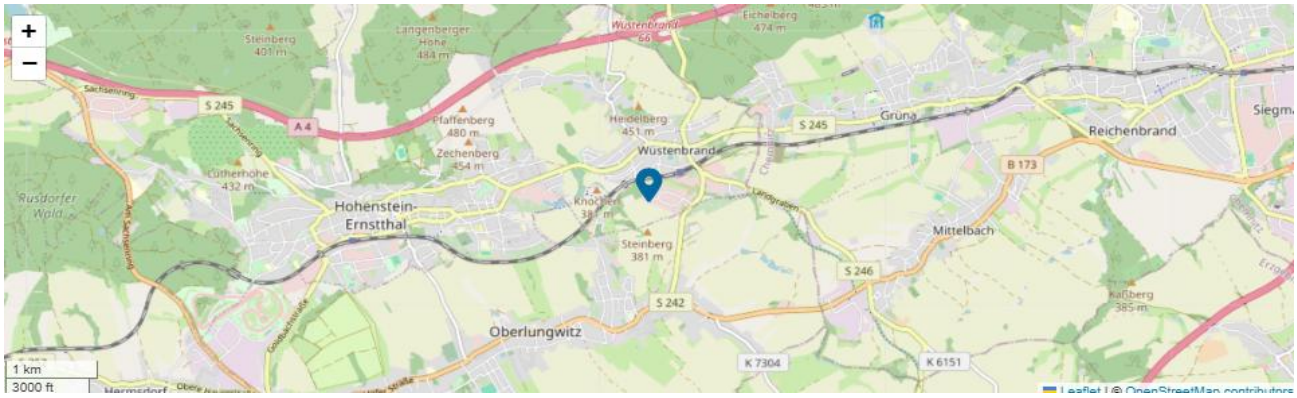


Abbildung 4: vollständige Messkette mit Datenausgabe

DOI 10.5162/iCCC2026/P8

3 ERGEBNISSE

Insgesamt belegen die Ergebnisse eine stabile und drift-freie Signalverarbeitung sowie eine hohe Messrobustheit unter Laborbedingungen und im realitätsnahen Feldversuch. Die Datenübertragung erfolgt zuverlässig, und das Gesamtsystem ist konsequent nach Ultra-Low-Power-Prinzipien ausgelegt.

Im Rahmen eines Proof of Concept wurde ein Messbereich von 0–30 % w/w bei stabiler Signalqualität validiert. Die Messungen zeigen eine robuste Performance sowohl im Tages- als auch im Nachtbetrieb.

Damit bestätigt das System seine grundsätzliche Eignung für die frühzeitige Waldbranddetektion. Perspektivisch sind umfangreiche Feldtests mit mehreren Sensorknoten sowie die technische Skalierung für den Serieneinsatz vorgesehen

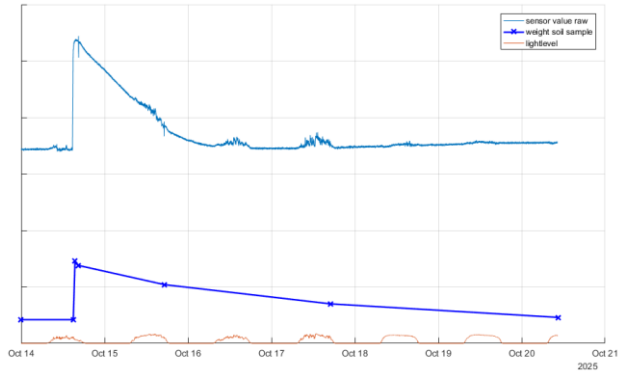


Abbildung 4: Korrelation zwischen Messwert und Bodenfeuchtigkeit

Abbildung 5: Sensor (USB Version)

Technische Daten des Sensors

Feature	Requirement
Messbereich	0% bis 30% w/w
Auflösung	mind 10 Pkt oder 3%
Genauigkeit	tbd
Elementarsensor	
Messfrequenz	< 1Hz
Ausgabe	USB Schnittstelle
Versorgung	über USB
Ausgabeformat	CSV
Arbeitsabstand	3 m
Messfläche	4 m^2
Umgebungs T°	-20°C +50°C
Luftfeuchtigkeit	0% 100%
Schutz	IP66
Stationärsensor	
Messfrequenz	2/Tag
Ausgabe	LTE
Versorgung	Batterie
Autonomie	1J
Ausgabeformat	JSON
Arbeitsabstand	2,5 m
Messfläche	4 m^2
Plattform	Theta
Umgebugs T°	-20°C +50°C
Luftfeuchtigkeit	0% 100%
Schutz	IP66

Abbildung 6: Datenblatt

4 Impact

5micron's IR- Soil- Moisture Messsystem adressiert eine der größten ökologischen und wirtschaftlichen Bedrohungen unserer Zeit: zunehmende Waldbrände infolge von Klimawandel und Trockenperioden. Das System ermöglicht eine frühzeitige Identifizierung von Risikozonen und reduziert dadurch den Bedarf an kostenintensiven Löschmaßnahmen sowie die Zerstörung von Wäldern, Infrastruktur und Gemeinden.

Zusätzlich verbessert es den Schutz der Biodiversität, unterstützt nachhaltige Waldbewirtschaftung und trägt zur Sicherheit der Bevölkerung in gefährdeten Regionen bei. Durch die einfache Installation und den autarken Betrieb können auch Regionen ohne bestehende Infrastruktur profitieren. Das mobile Einsatzkonzept eröffnet weitere Anwendungsfelder in der Landwirtschaft, im Katastrophenschutz und im Umweltmonitoring

DOI 10.5162/iCCC2026/P9

Miniaturisiertes Quantenlichtmodul zur Mikroplastikdetektion: MIR-Spektroskopie mittels verschränkter Photonen

Miniaturized Quantum Light Module for Microplastic Detection: MIR Spectroscopy Using Entangled Photons

Krista Bodenhagen[1], Gunnar Blume[1], Alexander Sahm[1], Felix Mauerhoff[1], Atta Ur Rehman Sherwani[2], Sven Ramelow[1], and Katrin Paschke[1]
[1]Ferdinand-Braun-Institut (FBH), Berlin, Deutschland
[2]Institut für Physik, Humboldt-Universität, Berlin, Deutschland
krista.bodenhagen@fbh-berlin.de

Kurzfassung

Spektroskopie im mittleren Infrarotbereich (MIR) ist aufgrund der hohen Eindringtiefe der MIR-Strahlung ein leistungsfähiges Werkzeug zur Analyse stark streuender Materialien wie (Mikro-)Plastik. Herkömmliche MIR-Quellen und -Detektoren sind jedoch in der Regel groß, teuer und technisch sehr anspruchsvoll. Daher wurde ein quanteninterferometrischer Aufbau basierend auf spontaner parametrischer Fluoreszenz (SPDC) in einem nichtlinearen ppKTP-Kristall zur Erzeugung verschränkter NIR-MIR-Photonenpaare realisiert. Nach der Erzeugung interagiert das MIR-Photon mit der Polymerprobe, während das verschränkte NIR-Photon einen Referenzweg durchläuft und von einem herkömmlichen Spektrometer detektiert wird. Aufgrund der Quantenverschränkung können Informationen über die Wechselwirkung des MIR-Photons mit der Probe allein durch die Detektion des NIR-Photons abgeleitet werden. Diese als „sensing with undetected photons" bezeichnete Technik ermöglicht MIR-Spektroskopie ohne direkte MIR-Detektion. Durch hochpräzise Mikrointegration und einen gefalteten Strahlengang für den Referenzarm konnte der optische Aufbau miniaturisiert und auf einer Fläche von nur 75 mm x 90 mm aufgebaut werden. In dieser Konfiguration wurden Photonenraten von über 10^9 pro Sekunde pro Watt Pumpleistung sowie eine Quanteninterferenz-Visibility von ≈ 30 % erzielt. Diese Ergebnisse unterstreichen die Eignung der entwickelten Module für mobile Anwendungen, wie etwa die vor-Ort Analyse von Mikroplastik in Abwasser, und stellen einen wichtigen Schritt in Richtung industrietauglicher Lösungen dar.

Abstract

Mid-infrared (MIR) spectroscopy is a powerful tool for analyzing highly scattering materials, such as (micro-)plastics, due to the high penetration depth of MIR radiation. However, traditional MIR sources and detectors tend to be large, expensive, and technically challenging. To overcome this limitation, a quantum interferometric setup based on spontaneous parametric down-conversion (SPDC) in a nonlinear ppKTP crystal was implemented to generate entangled NIR-MIR photon pairs. After generation, the MIR photon interacts with the polymer sample, while the entangled NIR counterpart travels through a reference path and is detected by a conventional spectrometer. Due to the quantum entanglement, information about the MIR photon's interaction with the sample can be inferred by detecting only the NIR photon. This technique, which is known as "sensing with undetected photons", enables MIR spectroscopy without the need for direct MIR detection. Using high-precision micro-integration and a folded beam path for the reference arm, the setup was miniaturized and built on an area measuring only 75 mm x 90 mm. In this configuration, photon rates exceeding 10^9 per second per watt of pump power and quantum interference visibility of ≈ 30 % were achieved. These results highlight the suitability of the developed micro modules for mobile and field-deployable applications, such as on-site microplastic analysis in wastewater, and marks a key step toward industry-ready solutions.

1 Introduction

Microplastics are ubiquitous and pose a growing threat to human health and the environment. However, detecting and identifying microplastic particles remains a major challenge due to their small size, the often low number of particles and the need for complex sample preparation. This process is further complicated by the high susceptibility of microplastics to contamination, such as from organic materials, and by the wide variety of existing polymer types, for which suitable reference materials are often lacking [1]. As a result, rapid and reliable on-site analysis, e.g. of wastewater samples, is both technically demanding and critically important, as it enables early detection and effective mitigation of potential health and safety risks. Mid-infrared (MIR) spectroscopy ($\lambda > 3\,\mu m$) has established itself as a powerful, non-destructive technique for the identification and characterization of polymer materials. However, conventional MIR sources and detectors are typically large, expensive, and technically complex, which significantly limits their applicability outside controlled laboratory environments [2]. This paper presents a novel approach to MIR spectroscopy based on the generation of entangled MIR-NIR (near-infrared) photon pairs within a

quantum interferometric setup. With this method, the sample is irradiated with MIR light, but detection of the MIR radiation is not required. Instead, the NIR photons, that are quantum mechanically entangled with the MIR photons, are detected and provide information about the sample. This innovative approach opens up a wide range of potential applications, especially for the analysis of materials at wavelengths for which no suitable detection mechanism currently exists.
The Ferdinand-Braun-Institut has succeeded in significantly miniaturizing this complex optical setup on a footprint measuring only 75 x 90 mm. This marks a decisive step towards mobile and user-friendly analysis of microplastic samples, opening up new possibilities for applications beyond traditional laboratory settings.

2 Concept and Design

This section describes the concept and optical design of the quantum module and its integration into the overall system. Afterwards, the assembly strategy of the micro-module is briefly outlined.

2.1 Interferometric Module

The interferometric setup is shown on the left-hand side in Figure 1. First, pump radiation is coupled into the module, collimated and focused into the center of a 30 mm long nonlinear crystal. In the periodically poled potassium titanyl phosphate (ppKTP) crystal, an incident pump photon is converted into an entangled signal–idler photon pair through spontaneous parametric down-conversion (SPDC). The frequencies of the generated photons are related through energy conservation, written as $\omega_P = \omega_S + \omega_I$ [2].

The module is designed to generate MIR radiation in the wavelength range from 3.2 to 3.6 µm, as many polymers exhibit characteristic absorption spectra in this spectral range. This results in wavelengths for pump and idler as shown in the Table below.

Radiation	Wavelength
Pump beam	720 nm
Signal beam	900-930 nm (NIR)
Idler beam	3.2-3.6 µm (MIR)

The resulting spectra for signal and idler appear to be very broad due to group-velocity matching of signal and idler realized by collinear type-0 quasi-phase matching (QPM) [2]. In order to meet the QPM conditions and thus achieve the desired signal and idler wavelengths and bandwidths, the crystal properties, in particular poling period and temperature, must be precisely matched to the properties of the pump laser, such as wavelength and polarization. A large bandwidth of the generated photon pairs is crucial for spectroscopy applications as it enables the investigation of samples over a wider spectral range [2].
A dichroic mirror behind the crystal separates MIR from NIR and pump beam, whereupon the MIR radiation leaves the module and interacts with the desired plastic sample. Meanwhile, signal and pump beam are propagating along a reference path within the module. The two lenses form a *4f* imaging system beginning at the middle of the crystal up to the end mirror of the reference path, minimizing chromatic aberrations that occur due to the large spectral bandwidth of the SPDC radiation and the wavelength difference between pump and signal. All the light is ultimately reflected back into the center of the crystal.
Here, interference occurs between backward propagating signal photons generated during the first pass and signal photons generated during the return pass through the crystal. Another dichroic mirror directs the NIR radiation out of the module and into a spectrometer. Due to wavelength-dependent phase differences, intensity modulations appear in the spectrum of the signal radiation, resulting in characteristic interference fringes.

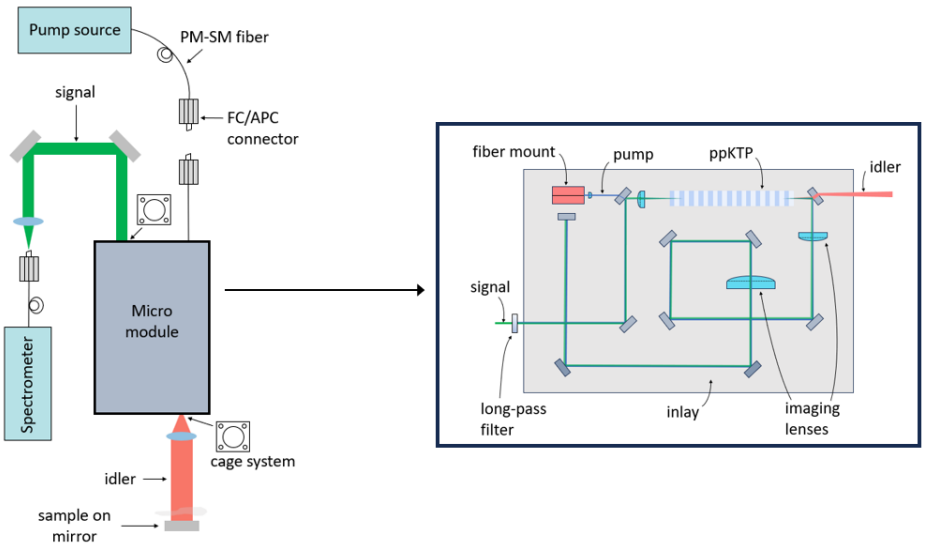


Figure 1 Optical layout of the micro module (*right*) and the overall experimental setup (*left*). Pump beam (*blue*) decays into signal (*green*) and idler (*red*) beam in the 30 mm long ppKTP crystal. The folded reference arm, starting at the second dichroic mirror behind the crystal, has a length of approx. 17 cm.

However, because the signal and idler photons are quantum-mechanically entangled, interference in the signal spectrum can only be observed if the idler path is precisely aligned and the two SPDC generation processes remain coherent and indistinguishable.
Consequently, if a sample that partially or completely absorbs MIR wavelengths is placed in the idler beam path, those altered spectral components are no longer indistinguishable and do not contribute to the quantum interference. In the NIR spectrum, this manifests as the disappearance of modulation in the corresponding spectral regions. The resulting changes in the signal spectrum can therefore be used to indirectly determine the absorption or transmission spectrum of the sample in the MIR range.

2.2 Overall Optical Setup

The integration of the module into the overall setup is shown on the right-hand side of Figure 1. The pump radiation is delivered via a polarization maintaining (PM) single mode (SM) fiber and an FC/APC fiber connector.
The MIR radiation propagates outside the module within a ~17 cm long standard 16 mm cage system, which contains a collimation lens. At the end of the cage system, an MIR-reflecting mirror is mounted, above which the polymer sample is positioned. Additionally, a focusing lens can be inserted to enable the investigation of smaller structures, such as microplastic particles. The NIR radiation also exits the module as a free-space beam and is coupled into the fiber leading to the spectrometer via two mirrors and a focusing lens mounted in another cage system. Moreover, another long-pass filter is positioned in the signal beam path to remove residual pump light. The high number of available degrees of freedom allows for continuous optimization of the coupling with respect to pump power, wavelength or idler alignment, enabling fine adjustment of the count rate in the spectrometer.
Another module was built with integrated FC/APC fiber connectors and internal fiber coupling of the signal radiation (Figure 2), enabling fast and efficient connection to both the pump source and the spectrometer. The adaption makes this module particularly user-friendly and allows for easy and compact integration into the overall setup.

2.3 Module Assembly

The module assembly was carried out using high-precision micro-integration technology and a combination of active and passive mounting techniques. In the active mounting approach, the optical components are positioned while the laser source is operating and subsequently fixed in place using UV-curable adhesive. In contrast, passive assembly relies on mechanical boundaries within the inlay that define the positioning of the optical components, without the use of an active laser source. While passive assembly enables highly time-efficient placement of the optical components, active assembly provides superior positioning accuracy. The fixation and alignment of the optical components are carried out using hexapods and vacuum tweezers, enabling highly precise positioning with step sizes as small as 100 nm.

Figure 2 Inside view of an SPDC module with integrated fiber connectors and internal fiber coupling of the signal radiation.

The inlay is fabricated from gold-plated copper-molybdenum, whereas the housing is made of anodized aluminum. The inlay has dimensions of 60 mm x 34 mm, and the overall module measures 75 mm x 90 mm. Furthermore, the module integrates a Peltier element for temperature control of the crystal, ensuring stable thermal conditions during operation.

3 Results

This section presents the basic functionality of the module, from quantum interference to plastic sample analysis.

3.1 Quantum Interference and Visibility

Figure 3 shows the measured NIR spectrum with and without quantum interference, e.g. with and without idler arm alignment, as well as the pure oscillation derived from the subtraction of the two spectra. The spectra were recorded at a pump wavelength of 720.8 nm, a crystal temperature of 40 °C, and an optical power of 15 mW within the module. The crystal used features a poling period of 23.375 µm. The spectrometer itself had an integration time of 50 ms and averaged over ten individual spectra. Overall, this results in a NIR photon count rate of > 10^9 per second per watt of pump power.

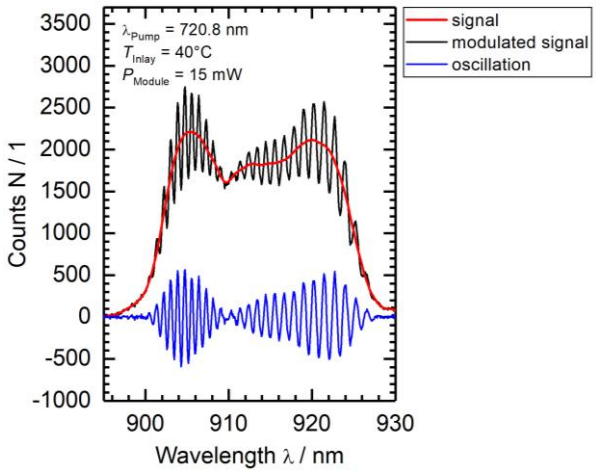


Figure 3 NIR spectrum with (*black*) and without (*red*) quantum interference fringes, as well as separated oscillation (*blue*).

DOI 10.5162/iCCC2026/P9

The visibility of the interference fringes is defined as $v = N_{max} - N_{min}/N_{max} + N_{min}$, which corresponds to the oscillation amplitude divided by the mean intensity. As can be seen in Figure 3, the visibility decreases towards the edges of the spectrum and reaches a local minimum around 910 nm. The drop in visibility in the center of the spectrum is attributed to intrinsic absorption within the crystal. Consequently, while a longer crystal increases the SPDC yield and thus the photon count rate, it also enhances self-absorption effects, leading to a loss of visibility. Considering only the two spectral regions with pronounced oscillations, an average visibility of approximately 30 % is obtained.
The alignment of the idler arm is crucial for achieving high visibility, as even slight angular deviations or changes in the idler path length reduce the indistinguishability of the SPDC processes. Consequently, the visibility decreases symmetrically with increasing misalignment around the point of maximum visibility [3].
Furthermore, the visibility, photon count rate, and spectral bandwidth strongly depend on the crystal temperature (see Figure 4). As the temperature decreases, the NIR spectrum becomes broader, while the visibility is reduced. This behavior can be attributed to a deviation from the QPM condition. When selecting the crystal temperature, it is essential to balance the generation of a broad spectrum with sufficiently high visibility to ensure that a meaningful transmission spectrum can be derived.

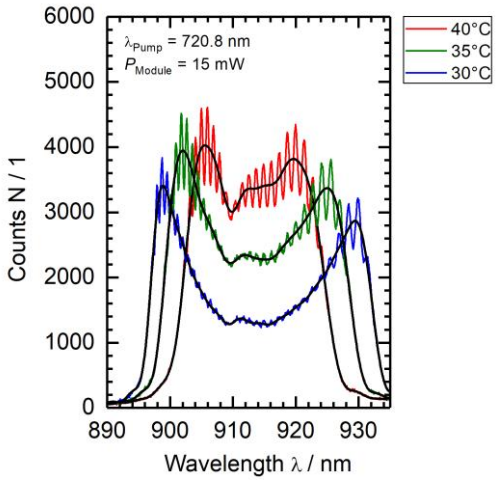


Figure 5 NIR spectrum for a crystal temperature of 30 °C (*blue*), 35 °C (*green*) and 40 °C (*red*).

3.2 Plastic Sample Identification

Figure 6 illustrates how the NIR spectrum with quantum interference changes when a sample is placed in the aligned MIR path. The modulation disappears entirely or partially in certain spectral regions. If both spectra, namely the modulated and the reference spectrum, are available, the envelopes of the interferograms can be obtained using a Hilbert transform, and from these the transmission curve of the sample in the MIR wavelength range can be determined. In this experiment, a plastic foil with a thickness of approximately 30 µm was mounted in the collimated beam path and analyzed. The spectrum shown was recorded with an integration time of 100 ms and an average of seven individual spectra.

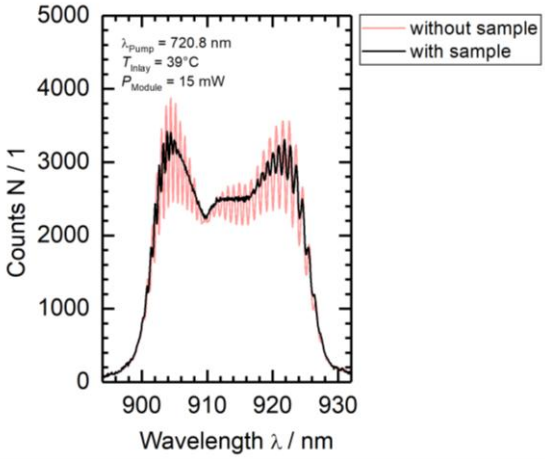


Figure 6 NIR spectrum with and without polymer sample placement in the idler arm.

4 Summary and Outlook

This paper presented the miniaturized implementation of a quantum interferometric setup for the generation of entangled NIR–MIR photon pairs. The compact design paves the way for transferring this technology from the laboratory to mobile systems, enabling applications such as wastewater analysis for microplastic detection. The basic operating principle of the module was demonstrated by examining a thin plastic foil placed in the aligned idler beam path.

Current efforts focus on the design of MIR-compatible microfluidic units, where microplastic particles can be aligned using hydrodynamic focusing under steady-flow conditions. A key challenge will be to maintain sufficiently high interference visibility despite the presence of absorbing glass substrates and water.

5 Acknowledgments

This work was funded under the project “SIM-QPla” by the German Federal Ministry of Education and Research (BMBF) under the grant number 13N15943. The authors would like to thank the project partners from the Humboldt-Universität zu Berlin and the University of Münster for their support.

6 Literature

[1] Kappacher C.; Badzoka J.; Huck C.: Validierung und QC in der Mikroplastikanalytik? WILEY Analytical Science, 2025.

[2] Vanselow A.; Kaufmann P.; Chrzanowski H.; Ramelow S.: Ultra-broadband SPDC for spectrally far separated photon pairs, Optical Society of America, 2019.

[3] Hildenstein P.; Blume, G; Sahm A. et al.: Efficient MIR sensing with undetected photons: High brightness, miniaturized SPDC module for real world applications. SPIE, 2025.

DOI 10.5162/iCCC2026/P10

Realisierung eines Demosystems für einen neuartigen Ansatz einer intelligenten und ultrakompakten Hyperspektralkamera

Realization of a demo system for an innovative approach of an intelligent and ultra-compact hyper spectral camera

Heinrich Engelke, Marcel Jongmanns, Friedrich Eckert, Fraunhofer Institut für Photonische Mikrosysteme, Dresden, Deutschland
heinrich.engelke@ipms.fraunhofer.de

Kurzfassung

Systeme für die hyperspektrale Bildgebung sind in vielen Anwendungen bereits für die Erfassung von Qualitätsmerkmalen oder für Sortieraufgaben etabliert. Dem Nutzen in der Anwendung steht immer der Aufwand für Installation und Betrieb des Systems gegenüber. Daher gibt es zahlreiche Bereiche, die aus Kostengründen bislang nicht adressiert werden können. Der technische Aufwand kann jedoch unter bestimmten Voraussetzungen signifikant reduziert werden, wenn beispielsweise die Objekte eine bekannte definierte Materialstruktur aufweisen. Ein neuer Ansatz setzt darauf, zunächst ein Bild zu erfassen und per KI- (künstliche Intelligenz) Algorithmus auszuwerten und dann einen ausgewählten Messbereich für die Spektralanalyse festzulegen und nur dort die Materialzusammensetzung zu erfassen. Hierdurch kann der Systemaufwand drastisch reduziert werden, ohne relevante Informationen einzubüßen.
Aufbauend auf dieses Konzept wurde ein Demonstrationssystem realisiert und in Betrieb genommen. Es zeigt die Systemfunktion in ausgewählten Beispielanwendungen. Erste vielversprechende Ergebnisse für die Sortierung von Textilien wurden erfasst. Darauf basierend, wird die Implementierung der Software für die erweiterte Nutzung der KI-Algorithmen ausgebaut.

Abstract

Hyper spectral imaging systems have entered numerous applications to access quality parameters or enable sorting tasks. The benefit of the application must be judged in comparison to the effort for installation and operation of the systems. Here, several applications cannot be addressed due to cost issues. Considering specific application scenarios there is a chance to reduce the effort significantly. If the application presents a well-known material structure, e.g. in textile recycling where an entire object is made from one material, simplified material estimation is sufficient to perform accurate sorting. A new approach starts with the acquisition of a visible image, evaluation using artificial intelligence which reveals a region of interest for the material analysis as a single measurement. The system effort is reduced without loss of relevant information.
Based on this concept, a demonstrator unit has been realized. It presents the system operation in selected applications. Promising results have been achieved for textile sorting. Based on the measurement results, the artificial intelligence algorithms are extended to more complex sorting tasks and further applications.

1 Einleitung

Konventionelle Hyperspektralkamerasysteme erfassen zu jedem Bildpunkt neben den Bildinformationen die chemische Zusammensetzung. Der Aufwand für die Erfassung der Daten für die Spektralanalyse und der Rechenaufwand für die Auswertung in Echtzeit sind sehr hoch. Für Anwendungen, in denen die Objekte vorwiegend aus einem Material bestehen, wird hierbei eine große Menge redundanter Informationen erzeugt. Das Konzept für einen intelligenten Ansatz, der aus dem Kamerabild einen Zielbereich für die spektrale Materialerfassung bestimmt und dort gezielt eine Messung durchführt, wurde erstellt (Abb. 1) und vorgestellt [1]. Die Umsetzung basiert auf der Auswahl des Zielpunkts mittels Ablenkung durch zwei MEMS Scannerspiegel mit quasistatischem Antrieb. Diese werden so ausgelenkt, dass ein auf Basis des aufgenommenen Bildes durch die KI ermittelter Messfleck spektralanalytisch vermessen wird.

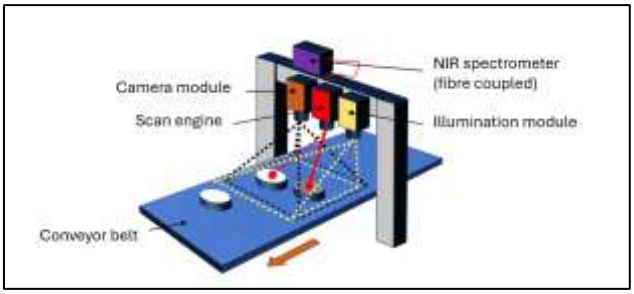


Bild 1 Schematische Darstellung des Systemkonzepts

Zur Messung sogenannter „real-time pattern“ und zur Demonstration und Evaluierung der Systemfunktion wurde

DOI 10.5162/iCCC2026/P10

ein erstes Messsystem aufgebaut (Abb. 2, links). Der Kamerakopf (Abb. 2 rechts) besteht aus einer Standardkamera und dem Spektrometersystem. Zusätzlich sind zwei Beleuchtungseinheiten montiert, die jeweils zugeschaltet werden können.
Die hiermit durchgeführten Messungen sind für die Entwicklung der KI-Algorithmen essenziell. In einer künftigen Variante werden dann die Ablenkeinheiten zur Positionsauswahl des Spektrometers und zum Mitführen der Bewegung des Förderbandes mittels KI-Algorithmus implementiert.

Bild 2 Demosystem der intelligenten Hyperspektralkamera(links), Spektrometer & Kamerasystem (rechts)

2 Intelligenter Ansatz unter Nutzung von MEMS-Technologie

Der neue Ansatz separiert Bildaufnahme und Spektralanalyse unter Nutzung von KI- (künstliche Intelligenz) Algorithmen. Die Spektralanalyse wird gezielt an einem ausgewählten repräsentativen Punkt durchgeführt. Hierdurch wird der Aufwand klassischer Systeme zur Erfassung und Auswertung von Spektraldaten jeden Ortspunkts, sowohl hardwareseitig als auch beim Rechenaufwand, drastisch reduziert.
Das Demonstrationssystem erfasst die Objekte im Bildbereich durch eine CMOS-Kamera mit 1920 x 1200 Pixeln (IDS GigE Vision). Diese nimmt ein Bildfeld von 40 x 40 cm² auf, das je nach Arbeitshöhe mittels telezentrischer Objektive mit 16 mm oder 25 mm Brennweite eingestellt wird. Die geometrische Auflösung des Bildbereichs ist im sub-mm Bereich. Die Bildrate beträgt standardmäßig 30 fps, kann aber auch auf 60 fps oder 100 fps erhöht werden. Die Beleuchtung erfolgt durch ringförmig angeordnete LED-Lichtquellen, sodass eine gute homogene Ausleuchtung sichergestellt und Schatten minimiert werden.
Für die NIR-Spektralanalyse wird ein neu entwickeltes Scanning Mirror Mikrospektrometer („SMMS“) mit verbessertem Signal – Rausch Abstand (*signal to noise ratio,* SNR) verwendet. Es erfasst wahlweise den nahinfraroten Spektralbereich von 950 nm bis 1900 nm oder von 1100 nm bis 2200 nm, jeweils mit 10 nm spektraler Auflösung. Aufgrund der hohen Arbeitsabstände ist ein gutes SNR für die korrekte Klassifizierung sehr wichtig. Besonders kritisch ist hierbei die gute und homogene Ausleuchtung mit geeigneter nahinfrarot (NIR) Strahlung. Zusätzlich kann eine höhere Integrationszeit durch Mitführen des Messflecks mit der Förderbandbewegung ermöglicht werden, die zudem eventuelle Bewegungsartefakte verhindert. Die hierfür eingesetzte Scannereinheit basiert auf zwei 1D Scannerspiegeln mit quasistatischem Antrieb, die mittels MEMS-Technologie (*micro electro mechanical system*s) in Silizium gefertigt werden. Der Auslenkbereich beträgt ±10 ° mechanisch, der optische Scanbereich ist somit 40 ° groß. Damit kann ein Bildbereich von ebenfalls 40 x 40 cm² für die Spektralanalyse adressiert werden. Der Durchmesser des Messflecks beträgt aktuell 5 cm, hierbei steht ein rauscharmes Signal ausreichender Höhe im Vordergrund. Für die primär adressierten Zielanwendungen wie Textilsortierung ist diese Geometrie vorteilhaft, sofern erforderlich können aber auch andere Messfleckdimensionen eingestellt werden.

3 Erzielte Messergebnisse

Unter Nutzung von aufgenommenen Bildern ausgewählter Beispielobjekte wurde die Bilderkennung angelernt. Zunächst stand die Erkennung der Objektgrenzen am Beispiel von Textilproben im Vordergrund. Anschließend wurde die Erkennung von Aufdrucken, Logos und anderen potenziellen Störfaktoren für die Spektralanalyse, wie z.B. Knöpfe oder Reißverschlüsse, vorgesehen (Abb. 3).

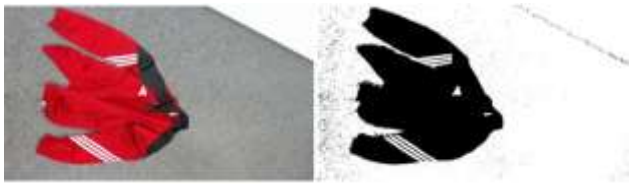


Bild 3 Erkennung von Textilproben, u.a. Außengrenzen (*borderline detection*), Aufdrucken, Logos

Für die Erfassung von NIR-Spektren wurde eine geeignete Beleuchtung mittel Halogenleuchten vorgesehen. Durch die Anordnung wird der Messbereich homogen und intensiv ausgeleuchtet. Nach einer kurzen Aufheizphase sind die Leuchten spektral stabil. Am Beispiel von Textilien aus Baumwolle, Polyester und Mischungen aus beiden Materialien wurden Spektren für die Klassifizierung aufgenommen (Abb. 4). Die Qualität der Signale lässt nicht nur eine qualitative Sortierung zu, sondern ermöglicht auch eine quantitative Aussage zur Zusammensetzung mit einer Auflösung von 5 % (absolut) oder besser.

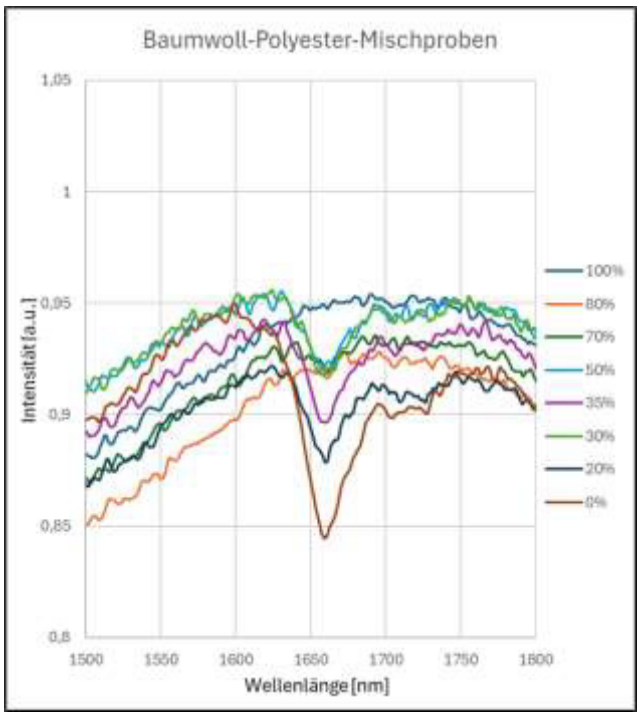


Bild 4 Spektren vom Textilproben

4 Schlussfolgerungen

Im Rahmen der laufenden Systementwicklung wurde mit der Fertigstellung der ersten Revision des Demosystems ein wichtiger Meilenstein erreicht. Das System ermöglicht die Erfassung von kontrastreichen gut ausgeleuchteten Bildern und nahinfrarot-Spektren. Es liefert die Messergebnisse zum Trainieren und Optimieren der KI-Algorithmen und erlaubt zudem die Bewertung der Komponenten wie Kamera, Objektiv, Scannereinheit, Spektrometer und Beleuchtungssysteme. Dieses Wissen ist speziell für die künftige Hochskalierung des Systemansatzes auf breitere Geometrien und schnelle Vorschubgeschwindigkeiten sehr relevant.

5 Ausblick

Die nächsten Schritte der weiteren Systementwicklung sind die Implementierung der KI-Algorithmen und darauf aufbauend die Realisierung einer Demo an einem laufenden Förderband. Hieran kann die Echtzeitfähigkeit von Software und Hardware erprobt werden. Neue Anwendungsbeispiele können effizient trainiert und in die Realanwendung überführt werden

Danksagung

Teile dieser Forschungsarbeiten wurden durch das BMFTR im Rahmen des Projekts “OASYS” unter Förderkennzeichen16ME0869 finanziell unterstützt

Literatur

[1] H. Grüger, J. Knobbe, T. Pügner, „Intelligente Hyperspektralkamera zur Erkennung und Bewertung von Objekten wie Textilien“, Posterbeitrag zur ICCC 2024
DOI 10.5162/iCCC2024/P23

[2] M. Knaack, H. Engelke, M. Jongmanns, S. Devi, I. Jablonski, “Smart and ultra-compact hyperspectral camera”, Posterbeitrag zum MST-Kongress 2025.

DOI 10.5162/iCCC2026/P11

Miniaturisierte Spektralanalysesysteme im Kontext von Lebensmittelqualität und Ernährungsoptimierung

Miniaturized systems for spectral analysis aiming at food quality and nutrition optimization

Heinrich Engelke, Tino Pügner, Fraunhofer Institut für Photonische Mikrosysteme (IPMS), Dresden, Deutschland
heinrich.engelke@ipms.fraunhofer.de

Kurzfassung

Es gibt umfangreiche Möglichkeiten sowohl die Lebensmittelqualität als auch Aspekte der individuellen Ernährung messtechnisch zu erfassen. Hieraus lassen sich Empfehlungen ableiten, die vom Kauf über Lagerung, Zubereitung bis hin zum Konsum den Anwender unterstützen. Neben allgemeinen Zielen wie der Reduzierung von Lebensmittelverschwendung („*food waste*“) können auch persönliche Ziel wie Gewichtsreduzierung, Fitnesssteigerung oder Muskelaufbau gefördert werden. Viele der aktuellen Ansätze haben jedoch Limitierungen in Bezug auf die Kosten, Anwendbarkeit oder den Aufwand für die Messungen. Hochintegrierte miniaturisierte Spektralanalysesystem bieten in zahlreichen Anwendungen eine Option für tragbare und einfach zu nutzende Systeme, die den vor Ort Einsatz ermöglichen.

Abstract

Several options can be applied to monitor food quality as well as aspects of individual nutrition. Knowledge of these can be used to advise on purchase, storage, preparation and consumption. Here personal aims can be supported besides healthy nutrition and reduction of food waste also more complex aims such as weight reduction, better fitness and muscle building. Existing systems mostly reveal limitations regarding cost, applicability and effort for the measurements. Highly integrated and miniaturized spectral analyzers offer options for portable and simple to use devices in numerous applications including mobile use.

1 Einleitung

Qualitativ hochwertige Lebensmittel und deren optimale Zusammenstellung sind der Schlüssel für die gesunde Ernährung. Neben der Ausgewogenheit von Ballaststoffen, Eiweiß, Kohlenhydraten und Fett sind auch die Berücksichtigung individueller Ziele relevant. Gewichtsreduzierung, Steigerung der Ausdauer oder Muskelaufbau können nur durch ein gezieltes Einstellen von Training und Ernährung bestmöglich erreicht werden. Hierzu ist nicht allein die Nahrungszuführung ausschlaggebend, sondern das, was individuell vom Körper aus der Nahrung heraus aufgenommen wird.

Die Spektralanalyse ist seit langer Zeit als Labormethode zum qualitativen und quantitativen Messen von Stoffzusammensetzungen etabliert. Erste Systeme stehen für den Feldeinsatz zur Verfügung [1, 2]. Von der Analyse des Ackerbodens, der gezielten Steuerung der Bewässerung, Begutachtung des Pflanzenwachstums und der Reife bis zur Ernte [3], anschließend Sortierung, Lagerung, Verkauf bis zur Zubereitung, Verzehr und Resteverwertung werden verschiedene Verfahren eingesetzt. Für die Zusammensetzungsanalyse spielt die nahinfrarot (NIR) Absorptionsspektroskopie eine tragende Rolle [4]. Probleme wie Schimmelbefall können mittels Fluoreszenzmessungen erkannt werden. Ebenso wird die spektrale Bildgebung („*hyper spectral imaging*“) im Anwendungskontext dort eingesetzt, wo der Nutzen den deutlich höheren gerätetechnischen Aufwand rechtfertigt.

In aktuellen Entwicklungen wird der vor-Ort Einsatz mit Echtzeitauswertung zunehmend wichtig. Ultrakompakte hochintegrierte Spektralanalysesysteme eröffnen neue Möglichkeiten solche Messungen durchzuführen und schnelle Entscheidungen, ohne den Umweg über die Laboranalyse zu ermöglichen.

2 Technologische Aspekte

Für die Umsetzung der Miniaturisierung werden von klassischen Laborgeräten bekannte Systemansätze weitgehend unverändert übernommen. Auch der grundlegende Ablauf der Spektralanalyse bleibt unverändert. Das Licht einer geeigneten Strahlungsquelle wird auf bzw. durch die Probe geleitet, nach der Wechselwirkung zwischen Photonen und Materie wird die Intensität zeit- oder ortsdiskret mittels Spektrometer erfasst und die spektrale Verteilung analysiert. Infolge materialspezifischer Absorption, Reflexion oder Fluoreszenz ist ein Stoffanalyse möglich. Die Spektralanalyse kann exemplarisch durch Monochromatoren, Interferometer als auch durch geeignete Spektralfilter erfolgen.

Anspruchsvolle laserspektroskopische Verfahren wie Raman-Spektroskopie, Kurzpuls- oder Mehrphotonenverfahren spielen derzeit in Anwendungskontext noch keine relevante Rolle, da der apparative Aufwand aktuell noch zu hoch ist.

DOI 10.5162/iCCC2026/P11

Ansätze für die Miniaturisierung werden insbesondere durch die Möglichkeiten der Mikrosystemtechnik vorangetrieben (Abb. 1: Smart-NIR). Hier kann beispielsweise das Beugungsgitter eines Monochromators in der beweglichen Silizium-Mikrostruktur eines MEMS-Bauelements (*micro electro mechanical systems*) realisiert werden oder ein Scannerspiegel wird zur Beleuchtung eines feststehenden Gitters genutzt.

Bild 1 Extrem miniaturisiertes NIR-Spektrometer

Neben MEMS-Bauelemente für aktive Komponenten in Spektrometern und Interferometersystemen tragen beispielsweise integrierte Lichtquellen, bevorzugt in kompakten Chipgehäusen wie Chip Scale Packages, zur weiteren Miniaturisierung der Gesamtsysteme bei (Abb. 2).

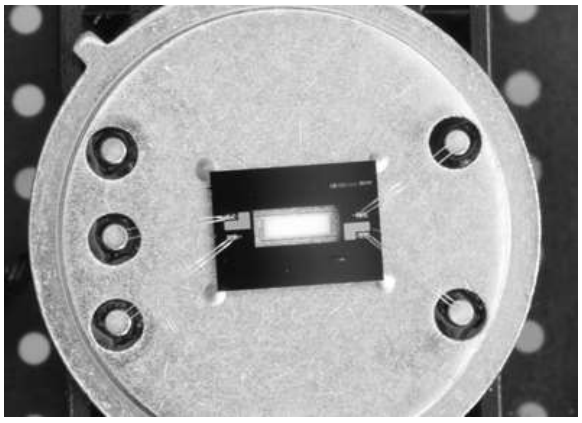


Bild 2 Infrarot-Lichtquelle im Chip Scale

3 Strategie für die Systementwicklung

Potenzielle Systementwicklungen profitieren von der Möglichkeit, zunächst Messungen mit regulärer Laborausstattung durchführen zu können (Abb. 3).

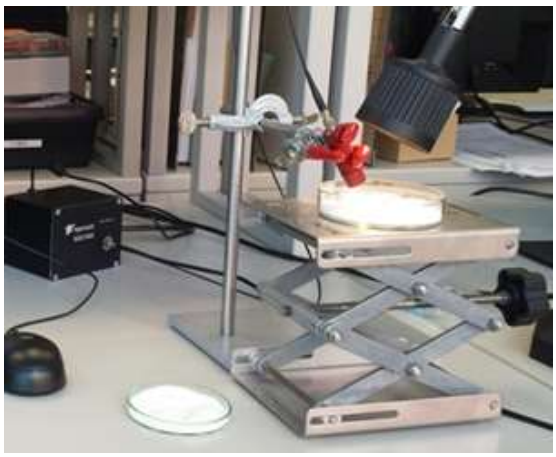


Bild 3 NIR-Absorptionsmessung in diffuser Reflexion (Laboraufbau)

Auf Basis der erfassten Ergebnisse (Abb. 4) kann eine Auswahl von Spektralbereichen bzw. Banden erfolgen, die für die Auslegung des miniaturisierten Messsystems notwendig sind. An dieser Stelle kann zudem entschieden werden, ob für die Erfassung der Spektraldaten ein Ansatz mit einem Monochromator notwendig ist oder ein polychromatischer Ansatz, im Sinne einer Anordnung aus mehreren Detektoren und Filtern passender Lage und Breite, günstiger umzusetzen ist.

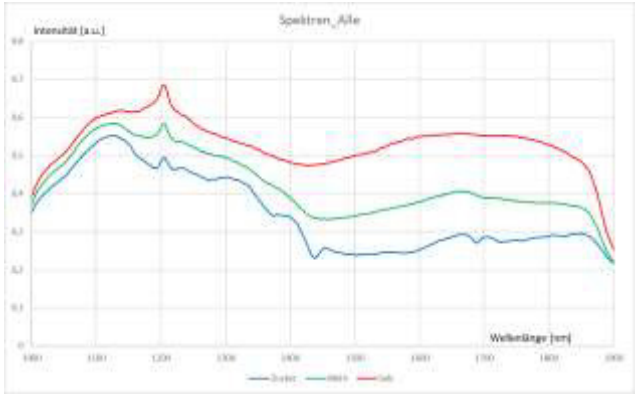


Bild 4 NIR-Absorptionsspektren von Zucker, Mehl und Salz – Lebensmittel der Gruppe „weiße Pulver“

Parallel kann mit der Software- und Modellentwicklung bereits begonnen werden, bevor das miniaturisierte System final fertiggestellt ist. Speziell beim bei der Nutzung von KI-Modellen (KI: künstliche Intelligenz), die zunächst aufwändig anhand großer Datensätze trainiert werden müssen, sind große Zeitvorteile erzielbar.

4 Systemintegration

Die Integration von ultrakompakten Optiksystemen erfordert neue Wege für den Systemzusammenbau. Speziell nichtaxiale Optiken, die in Monochromatoren und Interferometern Anwendung finden, sind komplex hinsichtlich der Montage und Justage der optischen Baugruppen. Der resultierende Herstellungsaufwand bringt Limitierungen für die Fertigung großer Stückzahlen mit sich und kann die

DOI 10.5162/iCCC2026/P11

Marktakzeptanz durch hohe Herstellungskosten einschränken. Ein neuer Ansatz für die Fertigung komplexer Optiksysteme ist die sogenannte „Faltmontage“ [5]. Hierbei werden mittels etablierter 2D Montageprozesse die Komponenten auf ein Substrat mit vorprozessierten Biegelinien assembliert. Durch Falten der Wände und des Deckels entlang der Biegelinien entsteht ein Körper mit dreidimensional angeordneten Funktionselementen und Strahlengang in seinem Inneren [6]. Wird eine ausreichende Genauigkeit erreicht, so können auch nichtaxiale Optiksysteme hocheffizient realisiert und große Stückzahlen kostengünstig gefertigt werden (Abb. 5).

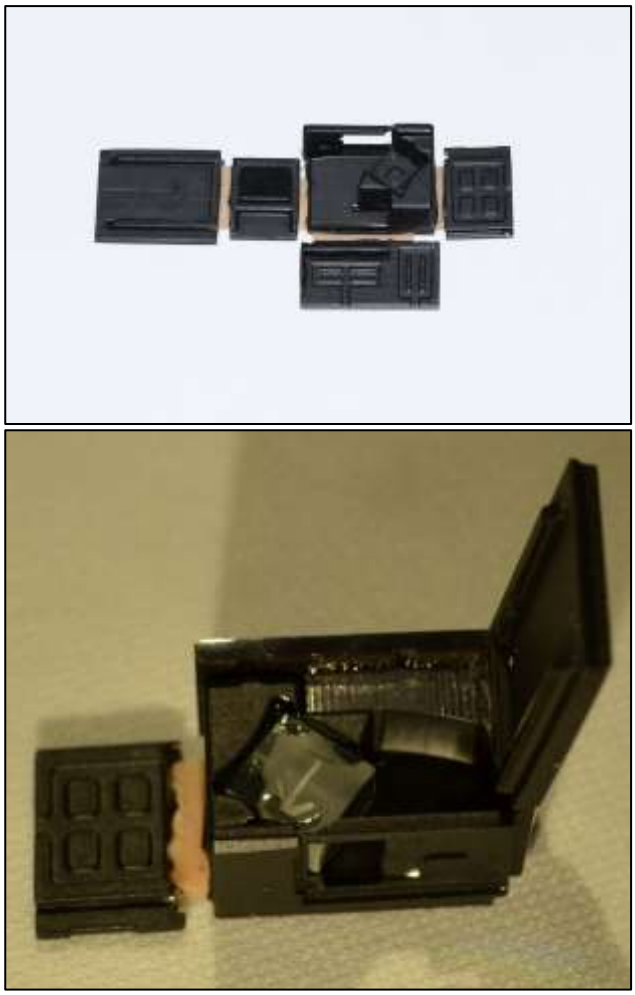


Bild 5 Integration eines Spektrometers mittels Faltmontage (oben unbestücktes Substrat, unten halbgefaltetes System)

Weiteres Potenzial ergibt sich aus den Möglichkeiten der additiven Fertigung. Hierdurch können bereits bei überschaubaren Stückzahlen ultrakompakte Systemkomponenten kostengünstig ermöglicht werden [7].

Die vorteilhafte Kombination aus Substraten, welche sich mit einer Auswahl an Fertigungsverfahren von kleinen bis großen Stückzahlen ökonomisch herstellen lassen, sowie etablierten Montagetechniken der Mikroelektronik, ergeben in Summe einem hervorragend stückzahl-skalierbaren Fertigungsansatz. Im Gegensatz zu anderen Ansätzen treten keine Technologiebrüche beim Übergang von kleinen zu großen Stückzahlen auf, was insbesondere mittelständischen Hersteller bevorteilt.

Aufbauend auf miniaturisierte Systemkomponenten wurde erste Handheldgeräte für den mobilen Einsatz realisiert.

5 Applikationsbeispiele

Die Nutzung hochkompakter Spektralanalysesysteme wurde bereits in ersten Anwendungen umgesetzt. Die nichtdispersive Infrarotmessung (NDIR) wird beispielsweise zur Erfassung des CO2 Gehalts in Luft eingesetzt. Hierbei reichen nur zwei Wellenlängen für eine quantitative Analyse aus.
Ähnliche Ansätze, jedoch im Bereich des nahinfraroten Spektralbereichs (NIR) ermöglichen die Messung der Feuchte von Böden für den Pflanzenanbau. Die bekannte Wasserbande bei 1450 nm wird zu einer Referenz, beispielsweise bei 1150 nm in Verhältnis gesetzt und ermöglicht so die Steuerung der Bewässerung.
Für die Messung von flüssigen Proben werden vorzugsweise Küvetten verwendet (Abb. 6). Die Messung in Transmission hat oftmals ein besseres Signal-Rausch-Verhältnis (SNR, *signal-to-noise ratio*) im Vergleich zur Messung in diffuser Reflexion. Je nach Absorptionsverhalten der Probe muss die Dicke der Küvette passend ausgewählt werden, um eine gute Aussteuerung des Spektrometers zu gewährleisten. Stark streuende Proben ergeben wenig Signal, hier ist unter Umständen die Messung über die Flüssigkeitsoberfläche in Reflexion vorteilhaft.

Bild 6 Messung an flüssigen Proben wie Milch, Olivenöl oder wässrigen Zuckerlösungen

Ein konkretes Beispiel ist die quantitative Messung des Zuckergehalts wässriger Lösungen. Hier wird eine Bande bei 1690 nm genutzt, die der Absorption von Haushaltszucker zugeordnet werden kann (Abb. 7).

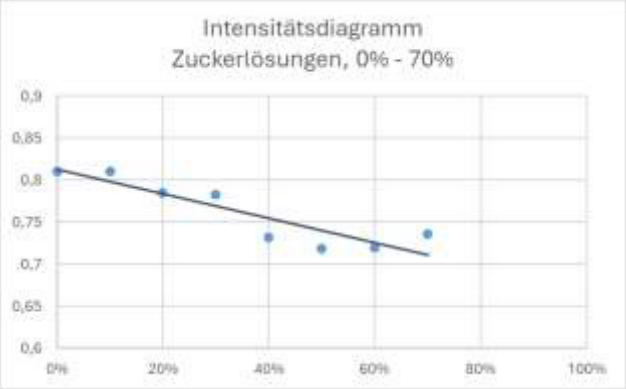


Bild 7 Spektrale Absorption in unterschiedlichen wässrigen Zuckerlösungen bei NIR-Wellenlänge von 1690 nm

Die Anwendung für komplexe Fragestellungen wie beispielsweise der Nährstoffaufnahme im menschlichen Körper, insbesondere Zucker bzw. Kohlenhydrate, Eiweiß und Fett, birgt die Möglichkeit, Ernährung anhand von Ist-Werten zu optimieren. Die Auswertung der Spektren ist jedoch komplex und erfordert innovative Algorithmen und Nutzung von künstlicher Intelligenz (KI).

6 Ausblick

Erste Anwendungen stehen kurz vor dem Beginn der Produktentwicklung. Neben der Hardwareintegration ist die Software- und Modellentwicklung wichtig. Selbstlernende Algorithmen unterstützen die Implementierung neuer Anwendungen. Neben tragbaren Systemen und sogenannten „Wearables“ steht für den Feldeinsatz die Nutzung von unbemannten Flugobjekten (Drohnen) und autonomen Robotern im Fokus neuer Entwicklungsprojekte.

Danksagung

Besonderer Dank gilt zahlreichen Nachwuchskräften, die im Rahmen ihres Engagements an den umfangreichen Labormessungen beteiligt waren.

Literatur

[1] T. Pügner, J. Knobbe, H. Grüger, “*Near-Infrared Grating Spectrometer for Mobile Phone Applications*”, Applied Spectroscopy 2016, Vol. 70(5), 734–745, (2016), doi: 10.1177/0003702816638277

[2] H. Grüger, J. Knobbe, T. Pügner, “*Improved scanning mirror micro spectrometer for material recognition in recycling applications*” SPIE Proceedings Volume 12899, MOEMS and Miniaturized Systems XXIII; 128990F (2024) https://doi.org/10.1117/12.3002165

[3] H. Grüger, J. Knobbe, T. Pügner, “*Multi- and hyperspectral imaging of plants and their properties by means of highly integrated and MEMS based systems*”, SPIE Proceedings Volume 12879, Photonic Technologies in Plant and Agricultural Science; 1287903 (2024) https://doi.org/10.1117/12.3002114.

[4] S.I. Hintschich, T. Pügner, J. Knobbe, J. Schröder, P. Reinig, H. Grüger, and H. Schenk, “*MEMS-based miniature near-infrared spectrometer for application in environmental and food monitoring for application in environmental and food monitoring*”, Proceedings of the 8th International Conference on Sensing Technology, ICST 2014.

[5] H. Grüger, J. Knobbe, T. Pügner, M. Leuckefeld, P. Reinig, S. Meyer, „*Concept for a new approach to realize complex optical systems in high volume*”, Proc. of SPIE Vol. 10448_26, (2017)
DOI: 10.1117/12.2279782.

[6] H. Grüger, J. Knobbe, T. Pügner, M. Schulz, S. Sdrenka, G. Ziegmann, “*Place and bend assembly: towards production of substrates in high volume technologies*” Proceedings Volume 12898, Advanced Fabrication Technologies for Micro/Nano Optics and Photonics XVII; 128980Q (2024)
https://doi.org/10.1117/12.3002152 PW 2024.

[7] H. Grüger, J. Knobbe, M.H. Sabiha, „*Investigation of mechanical and optical properties of 3D printed materials serving as substrate for place and bend assembly*“, SPIE Proceedings Volume 10931, MOEMS and Miniaturized Systems XVIII; 109310R (2019) https://doi.org/10.1117/12.2509279

DOI 10.5162/iCCC2026/P12

Design and Integration of Functional Blocks for Sub-THz Radar Sensor Systems up to 300 GHz

Batuhan Sutbas[1], Volkan Ertürk, Raqibul Hasan, Corrado Carta,
IHP–Leibniz Institute for High Performance Microelectronics, Frankfurt (Oder), Germany,
Paul Tschammer[2], Viktor Krozer, Wolfgang Heinrich,
Ferdinand-Braun-Institut für Höchstfrequenztechnik, Berlin, Germany,
Uwe Maaß[3], Usha Tamanna Yeasmin, Alexander Gäbler, Ivan Ndip,
Fraunhofer Institut für Zuverlässigkeit und Mikrointegration (IZM), Cottbus, Germany
[1]suetbas@ihp-microelectronics.com, [2]paul.tschammer@fbh-berlin.de, [3]uwe.maass@izm.fraunhofer.de

Abstract

Taking advantage of modern semiconductor processes and advanced integration technologies, the operation frequencies of fully integrated systems are continuously rising. The wide bandwidth available in the mm-wave and in the sub-THz frequency domains paves the way to various sensing applications. However, the design and integration challenges also increase with the complexity of such systems as well, often resulting in long development cycles. This paper highlights the functional block developments within the iCampus LausiTHz[1] project on radar sensor systems in D- (110-170GHz) and J-bands (220-325 GHz). By developing integration-ready RF blocks, transceiver ICs, and antennas as well as the capability of heterogenous integration, the project aims to enable faster development of such complex systems.

1 Introduction

The advantages of modern radar systems such as contactless measurement, operation in harsh environments, and all-electronic operation have led to new sensing applications. Exploiting modern semiconductor processes [1] and advanced packaging technologies, operating frequencies are continuously rising. Recent exploitation can be seen for in-vehicle sensing [2] and 3D imaging [3]. Operation of radar in the mm-wave and sub-THz spectrum, makes larger bandwidth available, increasing range resolution and reducing component size. However, with increasing frequencies, available output power and efficiency decrease while RF losses increase. In addition, using a larger bandwidth increases the severity of parasitic effects on RF signal paths and antennas. Therefore, output power and gain must be improved and integration losses minimized, while taking all electrical effects in the design of RF signal paths and antennas into account. There are already radar sensors and frontends available up to 240 GHz as system-on-chip integrating multi-channel transceiver and antennas [4]. While the on-chip integration of components allows a compact frontend, a compromise has to be found between the performance of transceiver and antenna, as well as the overall chip size. An alternative is the heterogeneous integration, combining various RFICs of different technologies and components in a single package [5]. In this paper, overview of the research conducted on the design of sub-THz transceiver and amplifier RFICs, on-chip antennas as well as the use of heterogeneous integration of SiGe and InP technologies for the realization of frontends and package-integrated antennas within the LausiTHz project is presented. The goal is to develop and evaluate technological building blocks usable for a faster development of future heterogeneously-integrated radar sensors.

2 Transmitter ICs

The 160 GHz transmitter (TX) assembly consists of a combination of a SiGe-based TX IC and an InP-based power amplifier (PA) chiplet flip-chip mounted onto it. This integration step is realized by a dedicated technology with micro bumps that offers bandwidth well beyond 200 GHz [5]. The SiGe IC is designed using the advanced IHP SG13G3 technology, which is based on [1], providing a transit frequency (f_T) of 510 GHz and a maximum oscillation frequency (f_{max}) of 600 GHz. It consists of a driver amplifier (DA) operating at Ka-band, a D-band frequency quadrupler, and a filter followed by a D-band DA feeding the InP PA chiplet, and a series of couplers and power detectors, before and after the InP PA for built-in monitoring of input and output power levels. The overall block diagram of the envisioned TX assembly is shown below in Fig. 1.

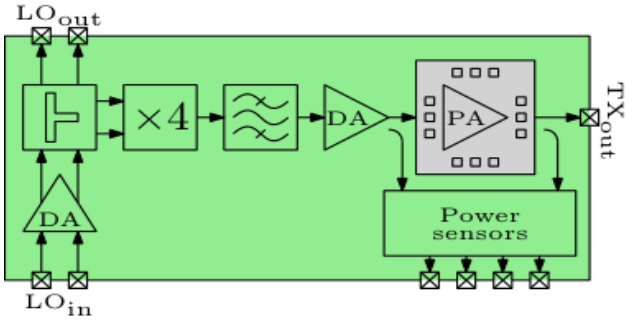


Figure 1 Block diagram of the D-band TX assembly with the SiGe TX IC (green) and InP PA chiplet (grey)

The SiGe TX IC is built based on the LO-scalable architecture [6], [7], allowing the number of TX channels to be expanded across the LO signal chain. This begins at the LO input port, where an external signal source provides a 39.5 GHz input signal. This signal is amplified and then

1 Funded by Federal Ministry of Research, Technology and Space of Germany (BMFTR) within project iCampus Cottbus.

split by a coupler into two paths. One path is routed to an output pad to facilitate LO scalability, while the other path feeds the on-chip frequency multiplier chain. Within the multiplication chain, the signal passes through a frequency quadrupler converting the signal to D-band, followed by a band-pass filter and a DA. To ensure integrity, the DA output is monitored by a power detector after a 20-dB coupler. Finally, the InP PA chiplet provides high-power amplification. The signal is transferred back to the SiGe IC at the output of the chiplet. The signal is sampled again with another 20-dB coupler to monitor the output power. The thru loss of the coupler is optimized to avoid degrading the TX output power. This configuration ensures that the signal is filtered, multiplied, and amplified to the necessary power levels while providing the readout mechanism required for real-time monitoring of the D-band output. A breakdown at building block level is given in Table I. The target bandwidth is at least 10 GHz.

Parameter→ Building Block↓	**IP1dB (dBm)**	**Gain/CG (dB)**	**PDC (mW)**
LO Driver Amplifier (DA)	−11	20	80
D-band quadrupler	5	−13.5	15
D-band DA	−3	11	20
Power detector	−15	-	1

Table I Expected parameters of the D-band TX IC

The D-band PA is realized as a 3-stage design based on InP HBTs with 700 nm emitter width. Each stage is based on a differential stacked common-base topology with respective baluns at the input and output. Below, Fig. 2 presents the layout of the RFIC. It is designed in an inverted microstrip architecture to avoid detuning by the flip-chip assembly. The simulated RF output power is approximately 17 dBm.

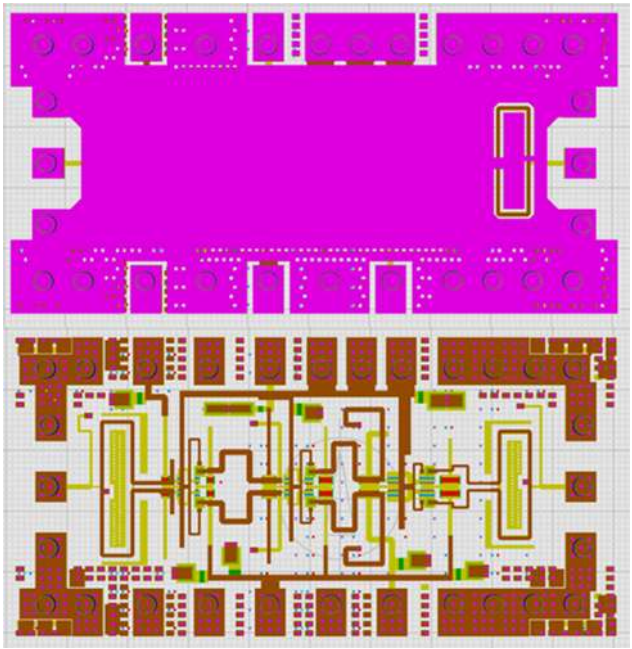


Figure 2 Layout of InP PA, top layer (top) and layer below top (bottom).

The J-band TX IC (Fig. 3 right), is also designed in the IHP SG13G3 technology, targeting an RF bandwidth of more than 50 GHz. An external LO source, provided by the RX IC counterpart, around 35 GHz drives the input. The LO signal undergoes a series of amplification stages in the Ka- and D-bands and frequency multiplication factors of 2 and 4. The doubler output at J-band is amplified by a 3-stage PA and radiated into air through on-chip antennas. The design is made fully differential for its desirable characteristics especially at sub-THz frequencies.

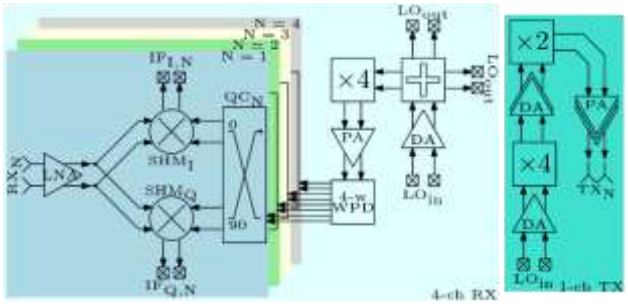


Figure 3 Block diagram of the J-band RX and TX ICs

The expected performance parameters of the functional blocks of the J-band TX are tabulated below (Table II). A power close to 15 dBm is expected at the antenna input with total simulated power consumption of 650 mW.

Parameter→ Building Block↓	**IP1dB (dBm)**	**Gain/CG (dB)**	**PDC (mW)**
LO Driver Amplifier (DA)	−4	12	10
D-band quadrupler	7	−10	15
D-band DA	−5	10	35
J-band doubler	4	0	20
J-band PA	−7	21	550

Table II Expected parameters of the J-band TX IC

3 Receiver ICs

The 160 GHz receiver (RX) IC is also designed in the IHP SG13G3 technology and features four channels. The RX IC mirrors the LO-scalable architecture of the TX, utilizing the same source. The IC consists of the same multiplication chain as the TX. Its output is converted back to differential signaling by means of a low-imbalance low-loss balun and amplified further by a D-band differential amplifier, feeding the LO ports of four quadrature RXs, after a 4-way power dividing structure. Each RX includes a quadrature coupler, two mixers, and an LNA with an integrated power divider (Fig. 4). On the RF signal end, each RX features an LNA providing 20 dB of simulated gain and only 5 dB of simulated noise figure. The incoming RF signal is split by a power divider to feed both quadrature mixers. The down-converted IF signals are routed to the output pads. This modular architecture allows for straightforward scaling of the number of elements in a RX array. Each RX chain is expected to have about 35 dB of overall conversion gain.

The J-band RX IC (Fig. 3 left), is designed in the IHP SG13G3 as well, targeting again an RF bandwidth of at

least 50 GHz. In contrast to the D-band version, the scalable RX IC at J-band has two LO outputs, one for cascading multiple RC RFICs, and the other for an LO source to the TX counterpart. The LO signal undergoes additional amplification after a multiplication chain similar to the TX to drive four quadrature RX channels. In contrast to the D-band RFIC, here the mixers are operated in a sub-harmonic manner, as it is more energy-efficient to obtain sufficient LO power. Expected performance of the functional blocks of the J-band RX are tabulated below in Table IV. The LO chain is expected to consume around 165 mW and each RX channel about 175 mW. An overall conversion gain of 22 dB is expected with a noise figure of only 11 dB.

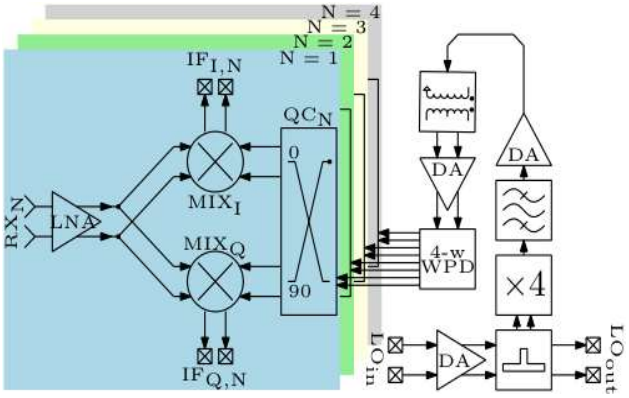


Figure 4 Block diagram of the D-band 4-channel RX IC

Parameter→ Building Block↓	IP1dB (dBm)	Gain/CG (dB)	PDC (mW)	NF (dB)
LO DA	−1	13	90	-
Quadrupler	5	−13.5	15	-
SE DA	−3	11	20	-
DE DA	−1	13	90	-
LNA	−19	20	35	5
Mixer	−12	20	5	8

Table III Expected parameters of the D-band RX IC

4 Antennas

4.1 On-Chip Antennas

On-chip antennas are the preferred choice for the J-band radar system. Integrating the antenna directly with the active circuit eliminates lossy transitions. The short wavelength at this frequency also allows miniaturization, making it easier to include the antenna on the chip [8].

Parameter→ Building Block↓	IP1dB (dBm)	Gain/CG (dB)	PDC (mW)	NF (dB)
LO DA	−4	12	10	-
Quadrupler	7	−10	15	-
DE DA	−4	12	70	-
LNA	−13	14	135	10
SHM	−9	11.5	20	17

Table IV Expected parameters of the J-band RX IC

The mentioned SiGe technology enables this integration, which offers excellent high-frequency performance. On the other hand, off-chip antennas for this frequency range suffer from severe signal loss and impedance mismatch due to package interconnects. The flip chip with solder bumps requires complex compensation networks to function properly while standard wire bonds generate high parasitic inductance levels [8], [9]. The design of on-chip antennas is also challenging, as the very thin metal layer in the standard back-end-of-line process results in a smaller spacing between the radiation element and the ground plane, which degrades antenna efficiency and bandwidth. Moreover, silicon has a high relative permittivity of around 11.9 and low resistivity, leading to significant energy dissipation and poor radiation efficiency. Due to the high dielectric constant in silicon, surface waves may be generated, which can distort the radiation pattern [10]. To overcome the challenges a micromachining post-processing technique known as localized backside etching (LBE) can be used. The lossy, high-permittivity silicon is removed under the radiating antenna element by LBE, effectively reducing conductive losses. This also suppresses the excitation of the surface wave. Therefore, an on-chip antenna with LBE significantly improves the radiation efficiency [9], [11].

4.2 On-Package Antennas

For the radar system at D-band, off-chip antenna options are evaluated, since the parasitic effects are not as severe. Various on-package antenna arrays at D-band (cf. Fig. 5) are analyzed based on an 127µm thick RF PCB substrate (Isola Astra MT77). Numerical simulations are performed to design the arrays and optimize the radiation performance.

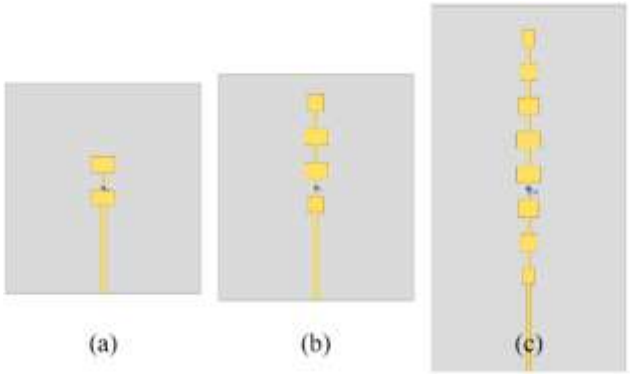


Figure 5 Design of 1×2, 1×4, and 1×8 antenna arrays

Antenna	1×2 Patch	1×4 Patch	1×8 Patch
Substrate	Isola Astra MT77 (ε_r 3, tanδ 0.004)		
Bandwidth	9.4 GHz	7.6 GHz	12.7 GHz
Return loss	40.9 dB	36.0 dB	60.2 dB
Gain	10.1 dBi	12.4 dBi	14.8 dBi
Beam width	17°	12°	7°
SLL	−11.1 dBc	−16.6 dBc	−15.2 dBc

Table V Simulated performance of various antenna arrays

Table V below gives a comparison of the simulated performance results for the patch antenna arrays. All designs show sufficient bandwidth and good matching in the frequency range of interest. With increasing array size, the antenna gain increases while the beam width decreases, as expected.

5 Integration

A PCB-based integration technology for the realization of the D-band frontend-antenna interface is investigated. For the IC interconnects, flip-chip and wire bond options are compared using numerical simulations. The 3D models used for EM simulations are shown in Fig. 6 and Fig. 7. The simulations results of flip-chip and wire bond transitions show sufficient RF performance in the frequency range of interest. The simulated transition and return losses are indicated in Fig. 8 and Fig. 9, respectively.

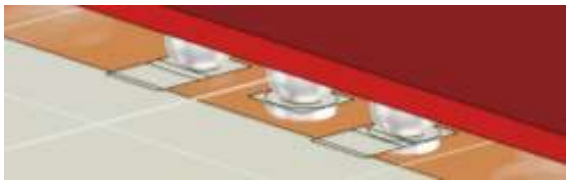


Figure 6 EM simulation model of Flip-chip transition

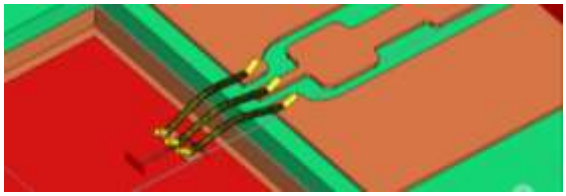


Figure 7 EM simulation model of wire bond transition

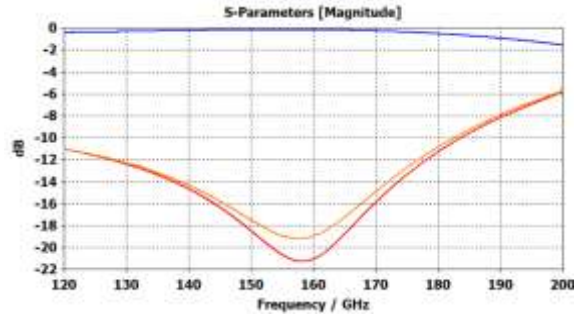


Figure 8 EM simulation results for flip-chip transition

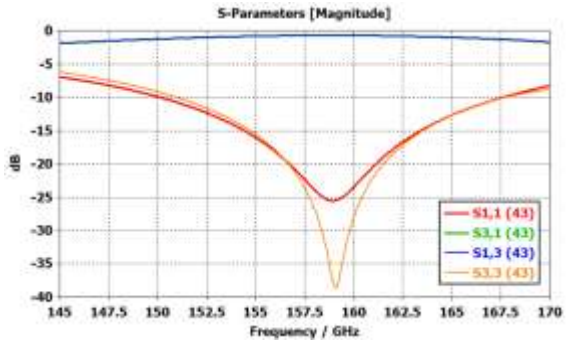


Figure 9 EM simulation results for wire bond transition

6 Summary and Outlook

Components and integration technologies for realizing sub-THz radar systems were investigated. The J-band radar system is planned to integrate at least three of each of the TX and RX ICs in the LO-scalable manner, where on-chip antennas are utilized. The D-band radar system is envisioned combining SiGe and InP technologies to take advantage of the capabilities of both domains. Expected results from the SiGe-based ICs, InP-based PAs, on-chip antennas, and PCB-based antenna arrays show promising simulated RF performance for the targeted sub-THz radar applications in the automotive and imaging domains.

7 References

[1] H. Rücker *et al.*, "SiGe-BiCMOS Technology with 500 GHz fT and 610 GHz fMAX," *2025 IEEE International Electron Devices Meeting*, San Francisco, CA, USA, 2025.
[2] A. Gharamohammadi, A. Khajepour and G. Shaker, "In-Vehicle Monitoring by Radar: A Review," in *IEEE Sensors Journal*, vol. 23, no. 21, pp. 25650-25672, Nov. 2023.
[3] M. Kahl *et al.*, "SiGe MIMO In-line Imager with 12x64 Elements for Real-time 3D Image Acquisition," 2*023 48th International Conference on Infrared, Millimeter, and Terahertz Waves*, Montreal, QC, Canada, 2023.
[4] M. H. Eissa *et al.*, "A Fully Integrated Modular 2×4 220–260 GHz Beam-Forming Transmitter and Receiver With 50 Gbps Wireless Transmission in SiGe:C BiCMOS," in *IEEE Transactions on Terahertz Science and Technology*, vol. 15, no. 5, pp. 805-820, Sept. 2025.
[5] M. Rausch *et al.*, "Broadband Hetero-Integration of InP Chiplets on SiGe BiCMOS for mm-Wave MMICs up to 325GHz," *2023 IEEE/MTT-S International Microwave Symposium*, San Diego, CA, USA, 2023, pp. 466-469.
[6] H. J. Ng, M. Kucharski, W. Ahmad and D. Kissinger, "Multi-Purpose Fully Differential 61- and 122-GHz Radar Transceivers for Scalable MIMO Sensor Platforms," in *IEEE Journal of Solid-State Circuits*, vol. 52, no. 9, pp. 2242-2255, Sept. 2017.
[7] W. A. Ahmad *et al.*, "Multimode W-Band and D-Band MIMO Scalable Radar Platform," in *IEEE Transactions on Microwave Theory and Techniques*, vol. 69, no. 1, pp. 1036-1047, Jan. 2021.
[8] A. Bhutani *et al.*, "Sub-THz Beam-Steering Antenna in Silicon Interposer Technology," *2024 54th European Microwave Conference*, Paris, France, 2024, pp. 808-811.
[9] H. J. Ng, R. Wang and D. Kissinger, "On-Chip Antennas in SiGe BiCMOS Technology: Challenges, State of the Art and Future Directions," *2018 Asia-Pacific Microwave Conference*, Kyoto, Japan, 2018, pp. 621-623.
[10] R. Hasan *et al.*, "On-Chip Slotted Patch Antenna at 226–244 GHz in SiGe Cu-Backend Technology," *2025 Asia-Pacific Microwave Conference*, Jeju, Korea, 2025.
[11] K. Schmalz, R. Wang, J. Borngräber, W. Debski, W. Winkler and C. Meliani, "245 GHz SiGe transmitter with integrated antenna and external PLL," *2013 IEEE MTT-S International Microwave Symposium Digest*, Seattle, WA, USA, 2013.

DOI 10.5162/iCCC2026/P13

Development of Bolus Sensor for Cattle Rumen Monitoring

Carsten Brockmann, Alexander Gäbler, Usha Tamanna Yeasmin, Sriram Charan Pasupuleti, Phillip Arnold, Uwe Maaß, Ivan Ndip, Fraunhofer Institut für Zuverlässigkeit und Mikrointegration (IZM), Cottbus, Germany
carsten.brockmann@izm.fraunhofer.de

Abstract

In order monitor the well-being of cows in grazing land scenario a rumen bolus monitoring system has been developed. It contains sensors for pH-, temperature and motion and is deployed inside the cow's rumen. The data is transmitted wirelessly from the bolus in the cow's rumen to anchor nodes using the LoRa protocol. The paper describers design of the sensor-, front-end-, wireless transceiver and power management electronic sub-systems as well as the integration into a bio-compatible housing along with the battery and an omni-directional antenna.

1 Introduction

Animal welfare is important for high-quality dairy products. Innovative technical systems that monitor the health of animals continuously can support farmers by determining the need for food or medication, detecting acute diseases and optimization of fertilization. There are already sensor systems for cows available for intracorporal application (Bolus sensors ([1], [2]) measuring activity and rumen secretion parameters. However, there is a lack of localization and real-time data connectivity to mobile networks (e.g. 5G).

In the framework of the project iCampus Smart Livestock Farming a networked multi-sensor system based on Bolus sensors that are deployed in the cow's rumen whose information is used for the early detection of abnormalities in the digestive tract and vital parameters, thus enabling targeted and species-appropriate treatment.

The Bolus features the following innovative technical solutions: Ultra-low power consumption, pH, temperature and motion measurement and wireless communication for system integration into a biocompatible Bolus-housing. This also inherits an optimized and miniaturized sub gigahertz antenna with omnidirectional gain which overcomes the challenges of attenuation and detuning inside the rumen. The Bolus transmits the sensor data wirelessly out of the cow rumen to distributed anchor nodes which allow a high coverage of even large pastures.

2 System Design

The monitoring system is designed for farmland animal keeping of cows on extended meadows where the cows can range freely. The autonomous bolus sensor system is deployed in the Rumen stomach of the cows. It transmits the sensor data wirelessly to anchor nodes distributed over the farmland using the LoRa ([3]) protocol. All anchor nodes are interconnected via a WiFi mesh network which is also used to upload the data to a server using a specialized 5G anchor node with additional 5G mobile network connectivity.

It is expected that the bolus monitoring system is operational for several years necessitating low-power system design. The rumen environment contains a high level of water creating high attenuation for radio signals especially at higher frequencies. To minimize the attenuation the LoRa standard operating at 433 MHz was chosen for the wireless link to the anchor nodes. In order to minimize the overall power budget a flexible wireless mesh network architecture was chosen allowing for topology optimization depending on the position of the cows.

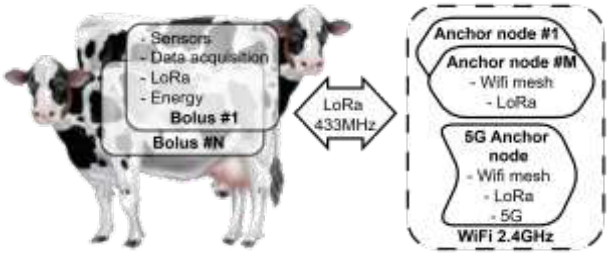


Figure 1 Smart Livestock system overview

3 Bolus Design

3.1 Bolus Construction

The sensor system is integrated into a cylindrical plastic housing fabricated of biocompatible and water-resistant Poly-oxy-methylene (POM) material. The lower part of the bolus contains the battery with a flex-PCB antenna wrapped around while the upper part contains the sensor front-end PCB (cf. Figure 2).

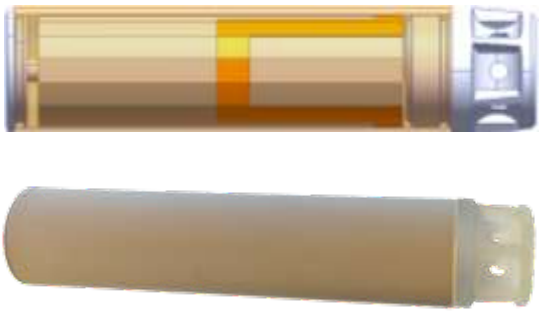


Figure 2 CAD design of bolus sensor (top) and fabricated bolus housing (bottom)

Table 1 gives the bolus outer dimension.

DOI 10.5162/iCCC2026/P13

Parameter	Value
Length	120 mm
Wall thickness	2,0 mm
Diameter	35 mm

Table 1 Bolus dimensions

3.2 Bolus Sensor Front-end PCB

The Bolus contains pH-, temperature- and motion-sensors connected to a low-power microcontroller with an integrated wireless transceiver (STM32W) implementing the LoRa protocol.
The electronics are realized using SMD technology on the front-end PCB. The antenna is connected externally to the front-end PCB using miniaturized coaxial RF connectors.
Figure 3shows the fabricated front-end PCB for set-up of the sensor system outside the bolus sensor.

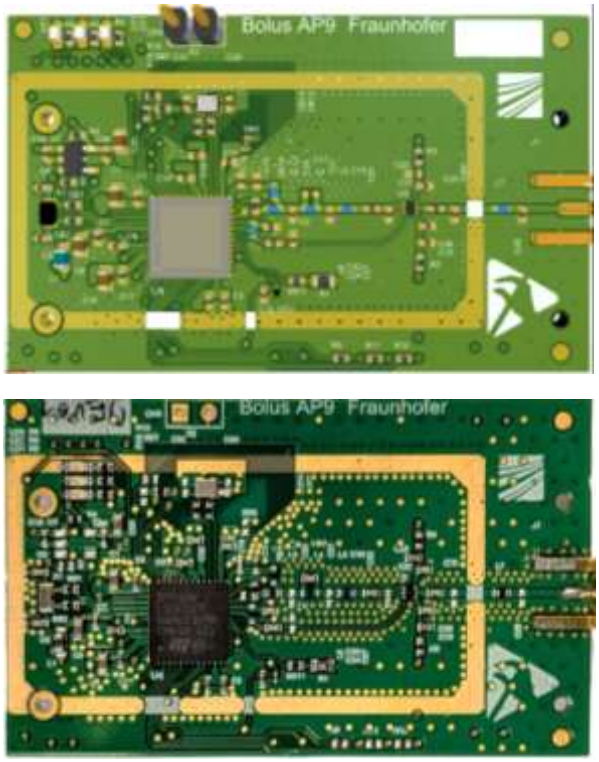


Figure 3 Bolus sensor front-end PCB. Top: layout, bottom: fabricated front-end PCB

3.3 Power Management

The Bolus is powered by a battery with a capacity of 19Ah. In order to meet the lifetime demands, the sensor system is operating intermittently.

4 Conformal Antenna Design

To transmit the collected data from bolus sensor to anchor node, the bolus application requires antenna which can be placed inside the bolus. For this bolus application, a flexible conformal minder antenna is designed on flexible substrate (dielectric constant ε_r= 3.4, and $\tan\delta = 0.002$) shown in Figure 4 so that it can be wrapped inside the bolus.

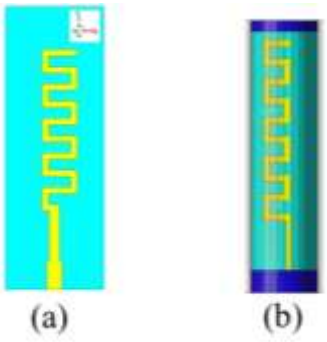


Figure 4 (a) Planar meander antenna (b) antenna wrapped inside bolus

The electromagnetic properties of the antenna strongly depend on the surrounding environments, causes impedance detuning and results in the loss of total efficiency. This antenna is designed based on considering the environment where the antenna will be embedded and place work to validate the antenna performance in terms of impedance matching and radiation. As this sensor bolus will be placed in the livestock rumen, the antenna is simulated inside a spherical homogeneous phantom of 150 mm radius and the cow rumen equivalent material properties (dielectric constant ε_r= 61, and conductivity σ= 1.77 S/m). This phantom helps to evaluate antenna performance specially the radiation pattern in the livestock body. Moreover, to realize the impact of the bolus electronics and battery, the bolus is filled with material with ε_r= 3.2, and $\tan\delta = 0.037$. In the antenna simulation, the thickness of the bolus is 0.2 mm, radius of the cylindrical bolus is 15 mm and length 110 mm, the material of the bolus is POM with ε_r= 3.8, and $\tan\delta = 0.0015$ are considered.

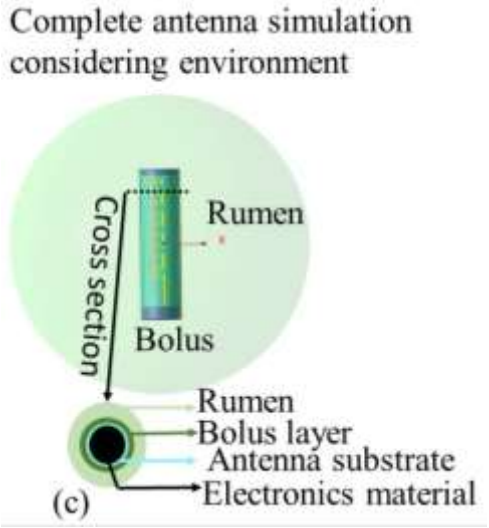


Figure 5 Antenna simulation considering environment

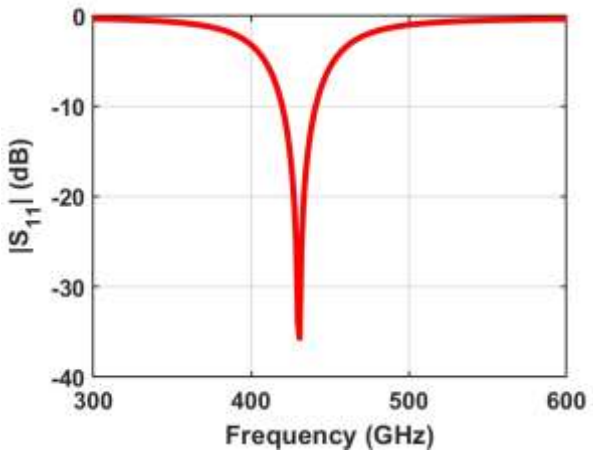


Figure 6 Reflection coefficient (S_11 in dB)

Figure 6 shows the simulation results of reflection coefficient (S_{11}) in dB for meander antenna array (Fig. 2(c)). The S_{11} curve shows a pronounced resonance at approximately 433 MHz, where S_{11} reaches a minimum value of about -21.6 dB. This deep notch in the reflection coefficient signifies strong impedance matching at the resonance frequency, with minimal reflected power and efficient power transfer into the structure. The achieved impedance bandwidth, defined by the -10 dB reflection criterion, is approximately 20 MHz, which exceeds the required bandwidth for the intended application.
In addition, the designed antenna exhibits an omnidirectional radiation pattern in 3D shown in Figure 7 with a simulated peak gain of -38 dBi. Despite the low gain, this radiation characteristic is sufficient to meet the required data transmission rate for the targeted short-range communication scenario.

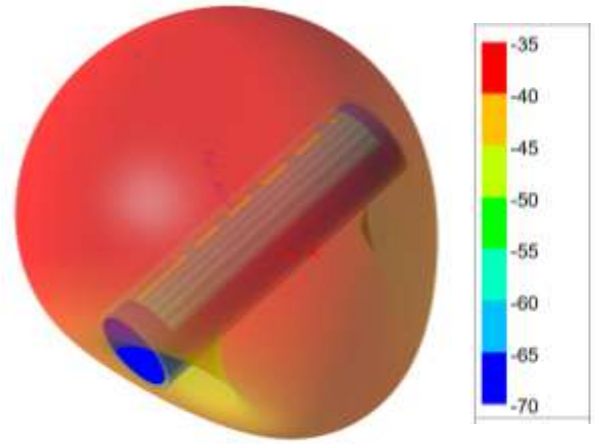


Figure 7 3D radiation pattern of antenna gain

5 Acknowledgement

The work was funded by the Germany research agency (BMFTR) in the framework of the project “iCampus 2 Smart Livestock”.

6 References

[1] Smaxtec System, https://smaxtec.com/de/
[2] Dropostix System, https://www.dropostix.com/de/produkt/
[3] LoRa alliance https://lora-alliance.org/

DOI 10.5162/iCCC2026/P14

Analyse industrieller Prozessdaten mit KI Agenten
AI Agentic Analysis of Industrial Process Data

Kai Gehrke, Sarah Spiesecke, Til Bartel, Ferdinand-Braun-Institut (FBH), Berlin, Germany, til.bartel@fbh-berlin.de

Kurzfassung

Wir zeigen, wie ein Large Language Model (LLM) allein durch Texteingaben komplexe Sensordaten einer Einzelwafer-Ätzanlage analysiert. Das Modell identifiziert Prozessläufe, normalisiert Zeitverläufe und erkennt Abweichungen mittels ±3σ-Hüllkurven. Die Analyse basiert auf MES- und Trace-Daten und erfolgt vollständig ohne Programmierung durch den Nutzer. Das LLM liefert innerhalb von Minuten visuell aufbereitete Ergebnisse und markiert Ausreißer. Der Ansatz demonstriert das Potenzial agentischer LLMs für schnelle, skalierbare Analysen in industriellen Anwendungen.

Abstract

We demonstrate how a commercially available Large Language Model (LLM) can execute complex data manipulation via simple text prompts. We prompt the LLM to analyze trace sensor data from a production equipment with the aim to establish a standard trace and assess possible process deviations. The LLM executes these tasks based on simple natural language input, demonstrating the increasingly agentic capabilities of LLMs. It is expected that this type of analysis will rapidly become a standard tool in R&D and industrial environments.

1 Motivation

Modern R&D and industrial environments are increasingly saturated with data. Thanks to automated systems and networked sensors, virtually every aspect of production and experimentation can be recorded and - in theory - analyzed. However, valorizing this data requires significant effort to clean and digest it. This typically demands domain knowledge and significant manual effort, often making it a bottleneck in data-driven optimization.

To address this challenge, we explore how a Large Language Model (LLM, the model), specifically ChatGPT, can act as an agentic analytical partner: a system capable of autonomously decomposing tasks, reasoning over data, and making purposeful decisions without explicit programming. In contrast, prior work using LLMs for data analysis [1,2] focuses on architectural adaptations and custom training to develop new machine learning (ML) algorithms, rather than leveraging the model's agentic capabilities.

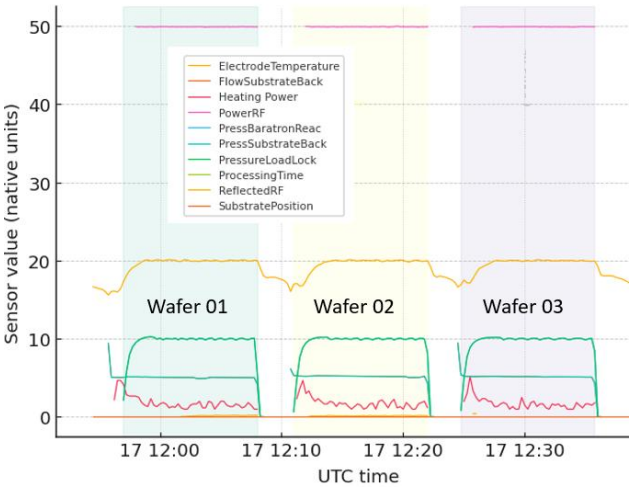


Fig. 1 ChatGPT o3-pro analyzes years of trace data and accurately identifies different sensor traces corresponding to single-wafer processing. In this example, a batch of four wafers is segmented correctly despite imprecise tracking timestamps. A visually compelling graph is generated.

2 Case Study: Analysis of trace data to identify process deviations

In this case study, we chose trace data from a reactive ion etching tool extensively used for processing wafers in our clean-room. Process lots comprise multiple wafers processed by the tool in individual runs. Metadata, including wafer names, recipe, and tracking times (for both lot and wafer-level events), is recorded in our manufacturing execution system (MES). The raw sensor time series, are stored via SECS-GEM protocol in an Influx data base.

Many different recipes are run on the tool, some of them of fixed length (etching by time), some of them of variable length (etching by endpoint detection). Contextual information from the MES (process start and end time stamps, recipe and wafer identification) have yet to be matched with the time series.

Browsing this heterogenous data to identify comparable runs and aligning them temporally is expected to hold important information regarding tool stability and health. An important use case is also helping to identify possible deviations in past and future runs.

2.1 Task description & prompting

The task can be divided in the following actions:

1) Ingest and understand both sources of data (Times series and MES context data)
2) Parse the meta data for a chosen run or recipe and resolve their time stamps
3) Select the trace data corresponding to these time stamps and identify typical traces, if needed, segment by wafer runs.

DOI 10.5162/iCCC2026/P14

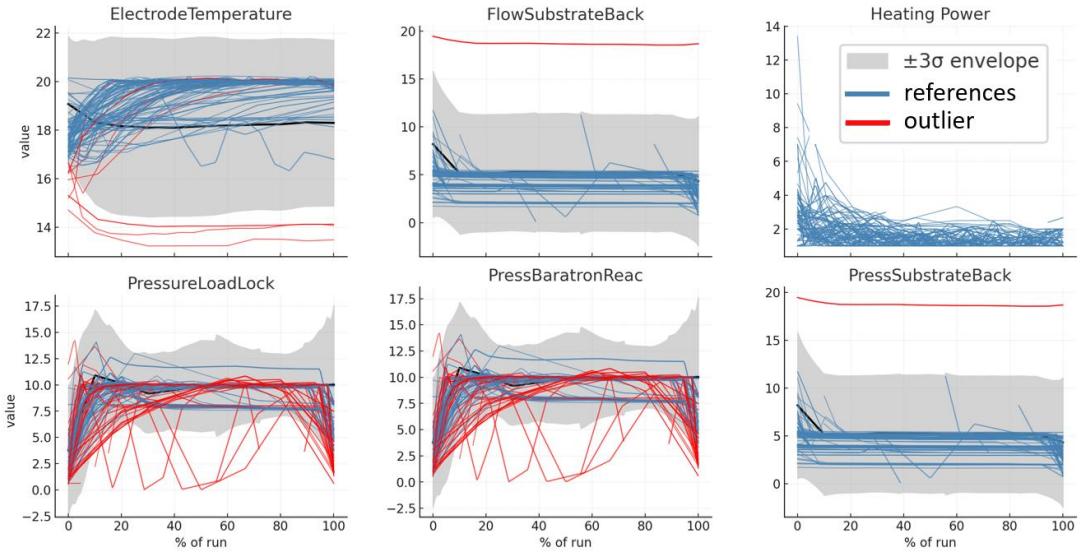


Fig. 2 The LLM has identified 157 runs processed with a given recipe on this particular machine. As the recipe has no fixed length, it was normalized and the run progress scaled from 0-100%. The 3-sigma envelope (grey shading) gives a good impression of the expected variations (healthy processes in blue). Processes that may have experienced deviations are shown in red.

4) Temporally align all selected runs, normalize to a common runtime if needed and compute a ±3σ envelope.
5) Generate plots and emphasize outliers

The entire workflow shall be driven solely through natural language prompts.

Approximately 5 MB of combined metadata and trace data were uploaded via the standard ChatGPT interface. The LLM (ChatGPT o3-pro) was prompted to perform the listed tasks.

2.2 Result and Discussion

Interpreting trace data, accurately mapping it to process runs, and establishing a reliable baseline are time-intensive tasks for human analysts. Furthermore, aligning these traces and normalizing their duration has always required manual intervention or custom algorithms. ChatGPT o3-pro, a model described as having “advanced reasoning capabilities”, remarkably excels at these tasks. The model processed two years’ worth of trace data, pre-filtered for a specific recipe, and successfully identified single-wafer runs in all cases, where the trace was within a 60-minute time window of the logged tracking times.

The model was able to distinguish and exclude incomplete processes, such as aborted runs or those with nonstandard profiles, and automatically filtered out data sets with unreliable tracking times, ensuring that only complete, valid traces were analyzed. The result is compellingly visualized in plots (Fig. 1 and Fig. 2). Runs that deviated beyond this envelope were automatically flagged as outliers and clearly annotated for further investigation.

Some of the tasks (run recognition, temporal alignment and skewing, plotting) were achieved in auditable Python code, that the model independently tuned both to the described tasks and the data analyzed.

The complete end-to-end analysis required less than 20 minutes, underscoring both the efficiency and power of agentic LLM-driven workflows—executed without custom coding, using only conversational prompts.

3 Outlook

We anticipate that agentic capabilities of LLMs - hosted on-premise or on trusted cloud-nodes and integrated with live access to databases and other organizational data sources (for instance using the model Context Protocol [3]) - will rapidly become standard tools in R&D and industrial environments. By combining historic and real-time data ingestion, these systems can continuously monitor process health, trigger alerts on emerging anomalies, and maybe at some point even propose corrective actions. The turnaround times for this type of analysis is reduced from days to minutes, and enables real time data-driven insights. Remaining challenges include reproducibility of results, which can often vary due to model stochasticity. These can be mitigated through advanced prompting strategies and the control of model parameters such as temperature.

As costs for running such models fall and tech-stack to integrate them mature, we expect such LLM-driven analytics platforms to transition from an experimental use towards a widespread tool, critical for advanced process control, continuous improvement and accelerating innovation cycles in manufacturing.

4 References

[1] X. Zhang, R. R. Chowdhury, R. K. Gupta, and J. Shang, “Large Language Models for Time Series: A Survey,” arXiv preprint arXiv:2402.01801, 2024.

[2] Y. Jiang et al., “Empowering Time Series Analysis with Large Language Models: A Survey,” arXiv preprint arXiv:2402.03182, 2024

[3] https://modelcontextprotocol.io/docs/getting-started/intro (retrieved on 31.07.2025)

DOI 10.5162/iCCC2026/P15

Untersuchungen der Schmierfettverteilung in Labyrinthdichtungen durch Kapazitätsmessung

Andreas Bürger, M.Eng., BTU Cottbus-Senftenberg, Senftenberg, Deutschland, andreas.buerger@b-tu.de
Prof. Dr.-Ing. habil. Sylvio Simon, BTU Cottbus-Senftenberg, Senftenberg, Deutschland, sylvio.simon@b-tu.de

Kurzfassung

Die Anwesenheit von Schmiermittel ist essentiell für einen dauerhaften verschleißarmen Betrieb von z.B. Getrieben, Lagern und Dichtungen. Während Qualitätssensoren zur in-situ Messung für Öl und Fließfett weit verbreitet sind, existiert kein vergleichbares Verfahren zur Erfassung der Anwesenheit und des Zustands von Schmierfett.
Bei der Untersuchung von Dichtungen und Lagern im Labor wurde bisher neben optischen Untersuchungen das benötigte Drehmoment als Indikator für die Anwesenheit und den Zustand des Schmierfetts benutzt. Durch die Messung der elektrischen Kapazität von Labyrinthdichtungen ist es möglich, weit mehr Informationen über die Schmierfettqualität und -verteilung und -bewegung in der Labyrinthdichtung zu erhalten als es z.B. mit einer Drehmomentmessung möglich ist.

Abstract

The presence of lubricant is essential for the long-term, low-wear operation of gearboxes, bearings and seals, for example. While quality sensors for in-situ measurement of oil and fluid grease are widely used, there is no comparable method for detecting the presence and condition of grease. In laboratory tests on seals and bearings, optical examinations have been used to detect the presence and condition of grease, along with the required torque. By measuring the electrical capacity of labyrinth seals, it is possible to obtain far more information about the grease quality and distribution in the labyrinth seal than is possible with torque measurement, for example.

1 Motivation

Zur Abdichtung von Wellendurchführungen durch Gehäuseöffnungen oder in Tragrollen von Gurtförderern kommen meist Labyrinthdichtungen in der Bauform senkrecht zur Achse zum Einsatz. Der Raum zwischen den Elementen der Labyrinthdichtung ist, zumindest im Neuzustand, vollständig mit Schmierfett gefüllt. Messungen zum Wirkprinzip und zum Verhalten des Schmierfetts erfolgten bisher im Labormaßstab durch Drehmomentmessung und durch optische Untersuchungen. Bei eigenen Untersuchungen an Labyrinthdichtungen zeigte sich, dass Messverfahren gefunden werden müssen, die es erlauben, das Verhalten des Schmierfetts in Labyrinthdichtungen auch im Realbetrieb zu untersuchen.

2 Untersuchungen in Labyrinthdichtungen

Von 2011 - 2018 erfolgten an der Universität Stuttgart umfangreiche Untersuchungen zur Wirkungsweise von Labyrinthdichtungen in der Bauform senkrecht zur Achse (fettgefüllte berührungsfreie Wellendichtungen, FBWD) [1][2][3]. Dabei wurde nachgewiesen, dass sich entgegen vorheriger Annahmen in der Labyrinthdichtung beim Betrieb ein Luftspalt öffnet, welcher die eigentliche Barriere für das Eindringen von Verschmutzungen in das zu schützende Lager darstellt. Der Nachweis erfolgte durch Kameraaufnahmen sowie durch Messung des Drehmoments. Es wurde gezeigt, dass das Absinken des Drehmoments auf nahe Null durch die Öffnung des Luftspaltes im Schmierfett, d.h. durch eine Änderung des Schmierfett-Luft-Gemisches, der Labyrinthdichtung verursacht wird. Das Drehmoment ist also ein direktes Indiz zur Öffnung des Luftspaltes.

Mit Drehmomentmessungen kann die Ursache der Drehmomentänderung in der Labyrinthdichtung jedoch nicht lokalisiert werden. Optische Untersuchungen wiederum bieten eine hohe Aussagekraft über die Vorgänge im Schmierfett, erfordern aber gute Lichtverhältnisse und ausreichend Platz für die Kamera.
Von den Autoren wurde an diese Arbeiten angeknüpft und weitere Untersuchungen an Labyrinthdichtungen durchgeführt [4]. Um die Vorgänge in der Labyrinthdichtung besser untersuchen und verstehen zu können, wurde der vertikale Teil der Labyrinthdichtung als Scheibe-Scheibe-System nachgebildet (Bild 1). Eine modifizierte Drehmaschine mit den Drehzahlen ab ca. 1 s^{-1} gefahren werden können, dient dazu, die Rotation des Scheibe-Scheibe-Systems realitätsnah nachzubilden.

Bild 1 Scheibe-Scheibe-System

DOI 10.5162/iCCC2026/P15

Der Raum zwischen beiden Ringen der Labyrinthdichtung d.h. zwischen den Teilen des Scheibe-Scheibe-Systems ist gewöhnlich vollständig mit Schmierfett gefüllt. Die üblicherweise für diese Anwendungen genutzten Schmierfette sind elektrisch nichtleitend. Damit liegt die Analogie zu einem elektrischen Kondensator nahe.

Untersuchungen der Dielektrizitätskonstanten von Schmierfetten ergaben Werte, die sich deutlich von der Dielektrizitätskonstante von Luft unterscheiden (Tabelle 1). Demnach sollte sich ein Eindringen von Luft in das Schmierfett durch ein Absinken der Kapazität bemerkbar machen.

Medium	ε_R
Luft (eigene Messung)	0,93
Schmierfett NLGI 2 (eigene Messung)	1,41
Schmierfett NLGI 3/4 (eigene Messung)	1,69
Luft [5]	1,00059
Wasser [5]	80,3

Tabelle 1 ausgewählte relative Dielektrizitätskonstanten

Damit die Teile des Scheibe-Scheibe-Systems als Kondensatorplatten genutzt werden können, mussten die aus Polyacryl bzw. Polyamid gefertigten Teile mit einer elektrisch leitenden Beschichtung versehen werden. Von der Vielzahl möglicher Technologien (z.B. Galvanisieren, Sputtern, Bedampfen, Bekleben mit einer Metallfolie) ergab lediglich der Auftrag mit einer eigentlich für elektromagnetische Abschirmzwecke bestimmten Farbe zufriedenstellende Ergebnisse bzgl. Haftung, Oberflächengüte und Reproduzierbarkeit.
Erste Messungen der Kapazität an einem einfachen Scheibe-Scheibe-System (Bild 1) erfolgten zunächst durch Messung der Entladezeit mit einem Arduino Nano 33. Diese Messungen ergaben reproduzierbar gute Ergebnisse und eine sehr gute Korrelation zwischen der Kapazitätsänderung und dem Absinken des Drehmoments. Letzteres wurde durch die Veränderung des Schmierfett-Luft-Gemischs durch Scherung verursacht (Bild 2).

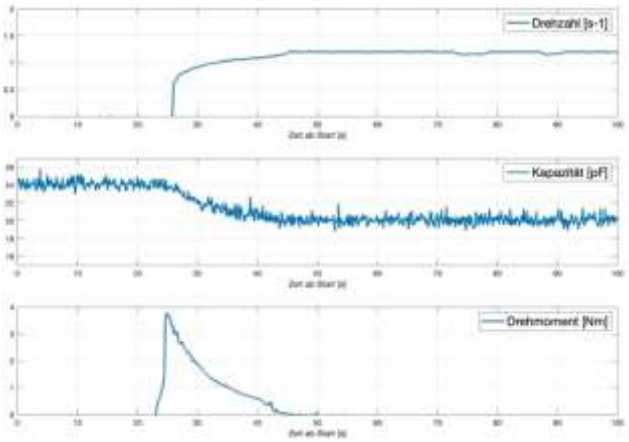


Bild 2 Messwerte für Drehmoment, Drehzahl und Kapazität am Scheibe-Scheibe-System (Bild 1)

Um die Anordnung als rotierendes System einsetzen zu können, wurde die rotierende Scheibe vollflächig, die stationäre Scheibe mit 6 Halbkreisringflächen beschichtet (Bild 3). Damit ergeben sich drei Kondensatoren, was eine Lokalisierung der Änderung der Schmierfetteigenschaften ermöglicht.

Bild 3 mit segmentierten Kondensatorflächen beschichtetes Scheibe-Scheibe-System

Die Nutzung des zur Kapazitätsmessung vorgesehenen IC FDC 1004 macht eine einfache und hochauflösende Messung der Kapazität bis in den Femto-Farad-Bereich möglich [6].
Bild 4 zeigt beispielhaft den Verlauf der Messwerte des Drehmomentes und der Kapazitäten C_{Innen} (innerer Ringkondensator), C_{Mitte} (mittlerer Ringkondensator) und $C_{Außen}$ (äußerer Ringkondensator) am Scheibe-Scheibe-System bei einer annähernd sprungförmigen Drehzahländerung von 0 s^{-1} auf ca. 26 s^{-1}.

Die Änderungen der Kapazitäten C_{Innen}, C_{Mitte} und $C_{Außen}$ die durch das Eindringen von Luft in das Schmierfett aufgrund drehzahlinduzierter Scherbewegungen verursacht wurden korrelieren deutlich mit Änderungen des Drehmoments. Markant sind auch periodische Änderungen der Kapazität - insbesondere im Bereich von C_{Innen} - welche durch Schmierfettbewegungen verursacht wurden. Diese Schmierfettbewegungen waren auch optisch erkennbar.
Ein Teil des Schmierfettes wird durch Fliehkräfte nach außen befördert. Es verbleibt jedoch immer ein Schmierfilm auf den Scheiben. Bei kleinen Drehzahlen ist der Einfluss der Fliehkräfte gering und der Großteil des Schmierfettes verbleibt zwischen den Scheiben und wird wulstartig zwischen den Scheiben bewegt.

Das Absinken des Drehmomentes ist zum Teil auf die Verringerung der Schmierfettmenge zwischen den Scheiben aufgrund der o.g. Wirkung der Fliehkräfte zurückzuführen. Bei kleinen Drehzahlen jedoch ist der Einfluss der Fliehkräfte gering und das Absinken des Drehmoments ist auf das Aufreißen der Lithiumbänder im Schmierfett zurückzuführen. Dieser Vorgang ist in der Literatur häufig beschrieben und durch z.B. REM-Aufnahmen belegt [7].

DOI 10.5162/iCCC2026/P15

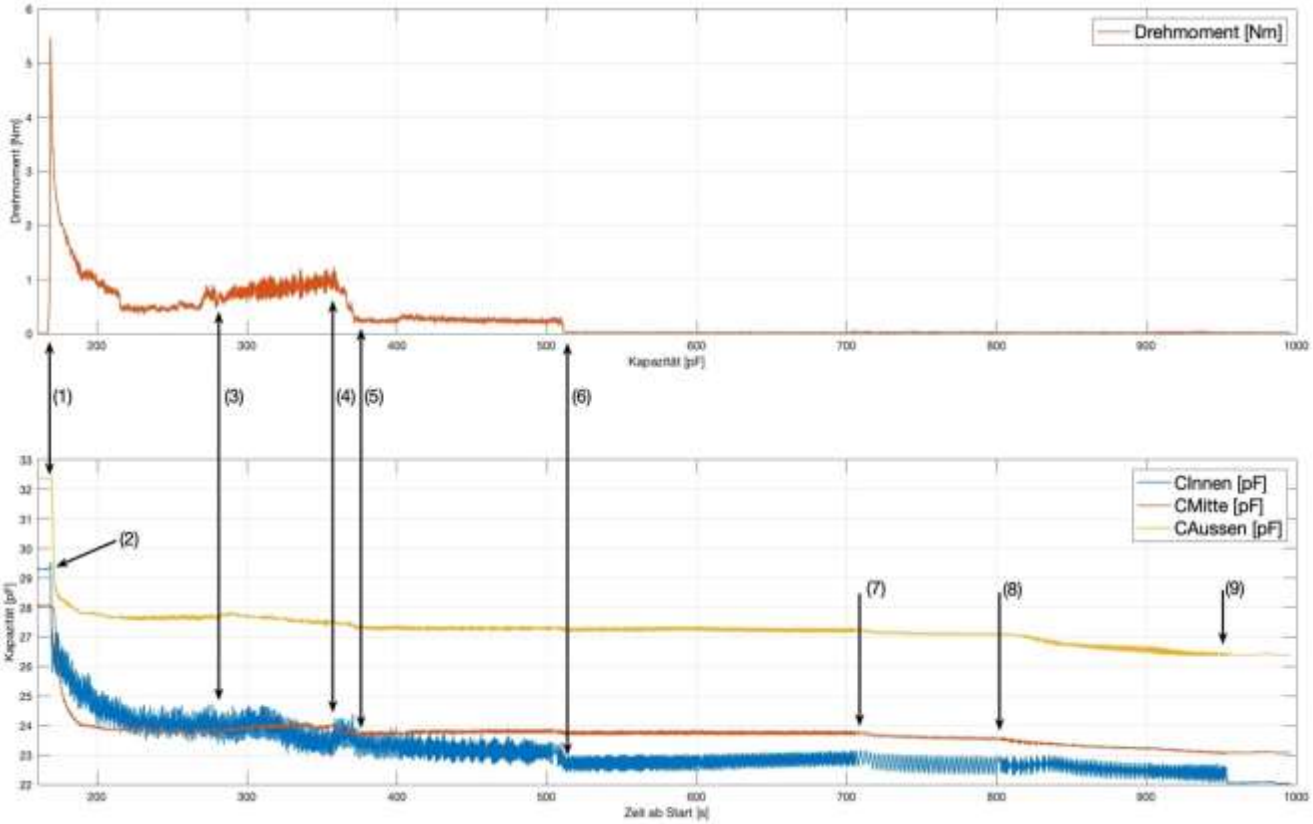


Bild 4 Messwerte für Drehmoment und Kapazitäten

Besonders markant sind folgende Vorgänge:

(1) Start des Versuchs, Anstieg des Drehmomentes
(2) steiles Sinken der Kapazität vermutlich infolge des Aufreißens der Schmierfettschicht und Eindringen von Luft
(3), (4), (5) Drehmomentänderungen, synchrone Änderungen der Kapazitäten
(6) Drehmomentänderung auf annähernd Null, weitere Kapazitätsänderungen
(7), (8) periodische Änderungen der Kapazitäten ohne messbare Drehmomentänderung
(9) Drehzahl 0 s^{-1}, Ende des Versuchs

3 Ausblick

Weitere Untersuchungen sollen an einer Nachbildung des waagerechten Teils der Labyrinthdichtung und an einer Kombination des waagerechten und senkrechten Teils erfolgen. Messungen mit unterschiedlichen Schmierfetten, unterschiedlichen Drehzahlen und Laufzeiten werden folgen.

In Zusammenarbeit mit dem Fraunhofer-Institut für Integrierte Schaltungen IIS wird derzeit eine Messvorrichtung entwickelt, mit der sich die Vorgänge im Schmierfett der Labyrinthdichtungen einer Tragrolle eines Gurtförderers während des Betriebes durch Messung der Kapazitäten untersuchen lassen. Parallel dazu werden mit dieser Messvorrichtung Dehnungsmessungen am Tragrollenmantel sowie Temperaturmessungen an verschiedenen Elementen der Tragrolle durchgeführt werden können. Die Messvorrichtung wird direkt an der Tragrolle montiert sein und während des Betriebes Messwerte in Echtzeit drahtlos übertragen.

4 Literatur

[1] Kümmel, J.U.: FKM-Vorhaben 287 „Fettgefüllte berührungsfreie Wellendichtungen mit Fettfüllung“, FKM-Forschungsberichte / Forschungskuratorium Maschinenbau; H. 311, Frankfurt/Main: 2011

[2] Kümmel, J.U.: Dissertation „Schmutzabdichtung mittels fettgefüllter berührungsfreier Wellendichtungen“, Berichte aus dem Institut für Maschinenelemente; Heft 171, Stuttgart: 2017, ISBN 978-3-936100-72-3

[3] Bosch, F.: Dissertation „Abdichtung trockener Stäube mit fettgefüllten berührungsfreien Wellendichtungen“, Berichte aus dem Institut für Maschinenelemente; Heft 177, Stuttgart: 2018, ISBN 978-3-936100-78-5

[4] Bürger, A., Simon, S., Hernschier, S., Moosburg, K.: „Untersuchungen an fettgefüllten, berührungsfreien Labyrinthdichtungen in Tragrollen“, 29. Fachtagung Schüttgutfördertechnik 2025, Magdeburg: Universitätsbibliothek Otto-von-Guericke-Universität, 2025, S.73, ISBN: 978-3-948749-55-2, http://dx.doi.org/10.25673/118679

[5] Dielektrizitätszahl-Liste, [Zugriff am 26. Mai 2025]. verfügbar unter https://www.vega.com/-/media/pdf-files/dielektrizitaetszahl-liste_de.pdf?rev=be01d5f43f94437c9de2b12c283e43e8

[6] Datasheet „FDC1004 4-channel 16-bit capacitance-to-digital converter with active shield driver for EMC protection“, [Zugriff am: 26. Mai 2025]. verfügbar unter: https://www.ti.com/de/lit/gpn/fdc1004

[7] Acar, N., Franco, J., Kuhn, E.: „On the shear-induced structural degradation of lubricating greases and associated activation energy: An experimental rheological study“, Tribology International 144, 2020, https://doi.org/10.1016/j.triboint.2019.106105

DOI 10.5162/iCCC2026/P16

KI-gestützte Messtechnik in der Siliziumwafer-Produktion

Dana Kaufmann, Gantner Instruments GmbH, Lauf an der Pegnitz, Deutschland
d.kaufmann@gantner-instruments.com

Abstract

Die Herstellung von Siliziumwafern für die Halbleiterindustrie erfordert höchste Präzision und Prozessstabilität. Insbesondere beim Schneiden (Dicing) und bei der Lithografie entstehen große Datenmengen, deren Analyse entscheidend für die Qualitätssicherung ist. Die Integration von Künstlicher Intelligenz (KI) in Messtechniksysteme ermöglicht eine Echtzeit-Überwachung kritischer Parameter wie Temperatur, Vibration und Schnittgeschwindigkeit. Dieser Beitrag untersucht die Lösungen von Gantner Instruments zur Implementierung von KI in die Wafer-Produktion. Vorgestellt werden Plattformen zur Datenerfassung und Steuerung (Q.series X), lokale Aggregation und Speicherung als Appliance (Q.core) und Datenerfassungslösungen in der Cloud für remote- und langzeit-Monitoring inklusive Analyse (GI.cloud), sowie deren Rolle bei der Erfassung und Analyse von Prozessdaten. Anwendungsbeispiele zeigen, wie KI-basierte Systeme die Ausbeute (Yield) erhöhen, Mikrorisse vermeiden und die Wartung von Dicing-Sägen optimieren. Abschließend wird ein Ausblick auf zukünftige Entwicklungen wie Edge-KI und digitale Zwillinge gegeben, die die Effizienz und Nachhaltigkeit der Halbleiterfertigung weiter steigern.

1 Status Quo

Die Halbleiterindustrie zählt zu den technologisch anspruchsvollsten Branchen weltweit und ist durch extrem hohe Anforderungen an Präzision, Reinheit und Prozessstabilität gekennzeichnet. Die Herstellung von Siliziumwafern bildet die Grundlage für nahezu alle modernen Mikroelektronikprodukte und umfasst eine Reihe komplexer Prozessschritte, darunter das Kristallziehen, das Schneiden der Wafer (Wafer-Dicing) sowie die Lithografie. Jeder dieser Schritte muss unter streng kontrollierten Bedingungen erfolgen, da selbst geringfügige Abweichungen in Temperatur, Vibration oder Schnittqualität zu erheblichen Qualitätsverlusten und hohen Ausschussraten führen können. Fehler in diesen Prozessen wirken sich nicht nur auf die Ausbeute (Yield) aus, sondern verursachen auch signifikante Kostensteigerungen in der Produktion.

Die Integration von Künstlicher Intelligenz (KI) in die Prozessüberwachung eröffnet neue Möglichkeiten zur Optimierung dieser hochsensiblen Fertigungsabläufe. KI-gestützte Systeme sind in der Lage, große Mengen an Prozessdaten in Echtzeit zu analysieren, Muster zu erkennen und Anomalien frühzeitig zu identifizieren. Dies ermöglicht eine vorausschauende Wartung (Predictive Maintenance) sowie die dynamische Anpassung von Prozessparametern, um die Qualität und Effizienz der Wafer-Produktion zu maximieren.

Gantner Instruments bietet hierfür hochpräzise Messsysteme, die sich durch ihre modulare Architektur und ihre Fähigkeit zur Integration von KI-Algorithmen auszeichnen. Die Plattformen Q.series X, GI.cloud und Q.core bilden die technologische Basis für die Erfassung und Verarbeitung kritischer Parameter wie Temperatur, Vibration und Schnittgeschwindigkeit. Durch die Kombination dieser Hardwarelösungen mit intelligenten Analyseverfahren können Hersteller nicht nur die Prozesssicherheit erhöhen, sondern auch die Produktivität und Nachhaltigkeit ihrer Fertigungslinien verbessern.

Bild 1 Entwicklung Sampling rates Quelle: Gantner Instruments GmbH

2 Anwendungen in der Waferproduktion

Die Integration von Künstlicher Intelligenz (KI) in die Wafer-Produktion bietet eine Reihe entscheidender Vorteile für die Prozesssicherheit und Effizienz. Ein zentrales Anwendungsfeld ist die vorausschauende Wartung (Predictive Maintenance) von Dicing-Sägen. Durch die kontinuierliche Analyse von Betriebsdaten können potenzielle Ausfälle frühzeitig erkannt und Wartungsmaßnahmen geplant werden, bevor es zu ungeplanten Stillständen kommt. Dies reduziert nicht nur die Ausfallzeiten, sondern senkt auch die Kosten für Notfallreparaturen.

Darüber hinaus ermöglicht die Echtzeit-Überwachung von Vibrationen eine präzise Kontrolle des Schneidpro-

DOI 10.5162/iCCC2026/P16

zesses. Vibrationen sind ein kritischer Faktor, da sie Mikrorisse im Silizium verursachen können, die die Qualität und Zuverlässigkeit der Wafer erheblich beeinträchtigen. Durch KI-gestützte Analyseverfahren lassen sich solche Anomalien sofort identifizieren und Gegenmaßnahmen einleiten.

Ein weiterer wichtiger Aspekt ist die Analyse von Temperaturprofilen während des Schneidens. Temperaturabweichungen können die Schnittqualität negativ beeinflussen und zu Materialspannungen führen. KI-basierte Systeme sind in der Lage, Temperaturdaten in Echtzeit zu verarbeiten und die Prozessparameter dynamisch anzupassen, um optimale Bedingungen sicherzustellen.

Bild 2 Leistung Messtechnik Quelle: Gantner Instruments GmbH

Insgesamt trägt die KI-gestützte Prozesskontrolle dazu bei, die Ausbeute (Yield) signifikant zu steigern und den Ausschuss zu reduzieren. Dies führt zu einer höheren Effizienz, geringeren Produktionskosten und einer verbesserten Nachhaltigkeit in der Halbleiterfertigung.

2.1 Predictive Maintenance

Die vorausschauende Wartung (Predictive Maintenance) stellt einen zentralen Anwendungsbereich der KI-Integration in der Halbleiterfertigung dar. Durch die kontinuierliche zeitsynchrone Datenerfassung und Analyse von Sensordaten in Echtzeit können KI-Algorithmen Abweichungen vom Sollzustand erkennen und potenzielle Ausfälle prognostizieren, bevor sie eintreten. Dies ermöglicht eine proaktive Planung von Wartungsmaßnahmen und verhindert ungeplante Stillstände, die in der hochsensiblen Wafer-Produktion mit erheblichen Kosten verbunden wären.

Die zugrunde liegenden Daten stammen aus einer Vielzahl von Sensoren, die kritische Parameter wie Vibrationen, Temperaturprofile und Schnittgeschwindigkeit erfassen. KI-Modelle verarbeiten diese Datenströme und identifizieren Muster, die auf bevorstehende Verschleißerscheinungen oder Fehlfunktionen hinweisen. Auf Basis dieser Erkenntnisse können Wartungsintervalle dynamisch angepasst und Ersatzteile rechtzeitig bereitgestellt werden.

Neben der Reduktion von Ausfallzeiten trägt Predictive Maintenance auch zur Verlängerung der Lebensdauer von Anlagen bei und senkt die Gesamtbetriebskosten. In der Halbleiterindustrie, in der Produktionsunterbrechungen besonders teuer sind, stellt diese Technologie einen entscheidenden Wettbewerbsvorteil dar.

2.2 Intelligente Datenanalyse

Traditionelle Messsysteme generieren enorme Mengen an Rohdaten, deren manuelle Auswertung zeitaufwendig und fehleranfällig ist. Die Integration von Künstlicher Intelligenz (KI) ermöglicht eine automatisierte und hochpräzise Analyse dieser Datenströme. KI-Algorithmen sind in der Lage, Daten in Echtzeit zu klassifizieren, Muster zu erkennen und Anomalien zu identifizieren, die auf potenzielle Prozessabweichungen oder Qualitätsprobleme hinweisen.

Durch den Einsatz von maschinellem Lernen können komplexe Zusammenhänge zwischen verschiedenen Prozessparametern erfasst werden, die mit klassischen statistischen Methoden nur schwer zu erkennen wären. Dies führt zu einer signifikanten Verbesserung der Entscheidungsgrundlagen für die Prozessoptimierung.

In der Halbleiterfertigung, insbesondere bei der Herstellung und dem Schneiden von Siliziumwafern, ist die intelligente Datenanalyse entscheidend für die Sicherstellung der Produktqualität. Sie ermöglicht die dynamische Anpassung von Prozessparametern wie Schnittgeschwindigkeit, Temperaturprofilen und Vibrationskontrolle, um die Ausbeute (Yield) zu maximieren und Ausschuss zu reduzieren.

Darüber hinaus unterstützt die KI-gestützte Analyse die Entwicklung digitaler Zwillinge, die als virtuelle Abbilder physischer Systeme dienen und für Simulationen sowie prädiktive Optimierungen genutzt werden können. Dies trägt nicht nur zur Effizienzsteigerung, sondern auch zur Nachhaltigkeit der Produktionsprozesse bei.

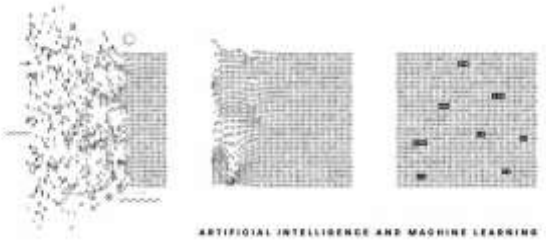


Bild 3 Data Analytics Quelle: Gantner Instruments GmbH

2.3 Effizienzsteigerung durch KI im Prozess

In der Halbleiterindustrie spielen Prüfstände und Testsysteme eine zentrale Rolle bei der Herstellung von Siliziumkernen und Wafern. Diese Systeme müssen unter streng

kontrollierten Bedingungen betrieben werden, um die geforderte Präzision und Qualität sicherzustellen. Die Integration von Künstlicher Intelligenz (KI) ermöglicht eine signifikante Effizienzsteigerung in diesen Prozessen.

KI-Algorithmen analysieren kontinuierlich die während der Tests generierten Daten und identifizieren Muster sowie Abweichungen in Echtzeit. Auf Basis dieser Analysen können die Testparameter dynamisch angepasst werden, um die Prozessbedingungen zu optimieren. Dies führt zu einer Verkürzung der Testdauer und der gesamten Herstellungszeit, ohne die Qualität zu beeinträchtigen.

Darüber hinaus erlaubt die KI-gestützte Steuerung eine adaptive Prozessführung, die auf Veränderungen in Materialeigenschaften oder Umgebungsbedingungen reagiert. Dies ist besonders relevant in Reinraum-Umgebungen, in denen selbst kleinste Schwankungen erhebliche Auswirkungen auf die Produktqualität haben können. Durch die Kombination von präziser Messtechnik und intelligenter Datenanalyse wird nicht nur die Produktivität gesteigert, sondern auch die Ausschussrate reduziert, was zu einer höheren Ausbeute (Yield) und geringeren Produktionskosten führt.

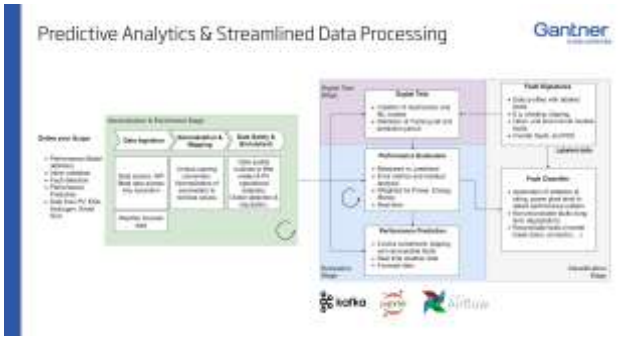


Bild 4 Predictive Analytics Quelle: Gantner Instruments GmbH

2.4 Energieeffizienz und intelligente Steuerung in der Waferproduktion

Die Herstellung von Siliziumwafern ist ein energieintensiver Prozess, der mehrere thermische und mechanische Bearbeitungsschritte umfasst, darunter Kristallziehen, Wafer-Dicing und Lithografie. Diese Prozesse erfordern eine präzise Steuerung von Temperatur, Vibration und Schnittgeschwindigkeit, um die geforderte Qualität sicherzustellen. Gleichzeitig steigt der Druck, den Energieverbrauch zu senken und die Ressourceneffizienz zu erhöhen.

KI-gestützte Systeme bieten hier entscheidende Vorteile. Durch die kontinuierliche Analyse von Sensordaten in Echtzeit können Algorithmen nicht nur Anomalien erkennen, sondern auch die Energiezufuhr dynamisch anpassen. Beispielsweise lässt sich die Heizleistung in Kristallziehöfen optimieren, um den Energieverbrauch zu reduzieren, ohne die Materialqualität zu beeinträchtigen. Ähnlich können die Antriebsparameter von Dicing-Sägen so gesteuert werden, dass unnötige Lastspitzen vermieden werden.

Darüber hinaus ermöglicht die Integration von KI in die Produktionssteuerung die Entwicklung intelligenter Energiemanagementstrategien, die vergleichbar mit Smart-Grid-Konzepten sind. Durch die Vernetzung mehrerer Produktionslinien können Lastprofile analysiert und Energieflüsse optimiert werden. Dies trägt nicht nur zur Senkung der Betriebskosten bei, sondern unterstützt auch die Einhaltung von Nachhaltigkeitszielen und regulatorischen Vorgaben zur CO_2-Reduktion.

Langfristig wird die Kombination aus präziser Messtechnik, KI-gestützter Prozessoptimierung und intelligentem Energiemanagement die Grundlage für eine ressourcenschonende und wirtschaftlich effiziente Waferproduktion bilden.

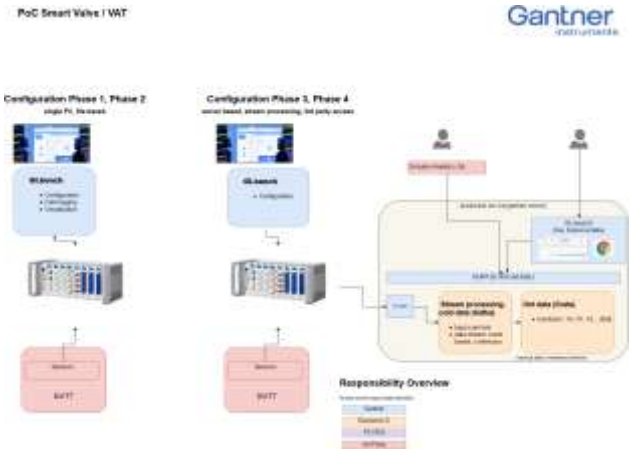


Bild 5 PoC Smart Valve Quelle: Gantner Instruments GmbH

2.5 Forschung und Entwicklung

Die Anwendung von Künstlicher Intelligenz (KI) in der Forschung und Entwicklung der Halbleiterindustrie eröffnet neue Möglichkeiten zur Analyse und Optimierung komplexer Prozesse. Insbesondere bei Strukturtests, die für die Bewertung der mechanischen und thermischen Eigenschaften von Materialien und Komponenten entscheidend sind, bietet KI erhebliche Vorteile.

Durch die Kombination von hochpräzisen Messdaten (Rohdaten plus Metadaten) mit **digitalen Zwillingen** können Ingenieure virtuelle Abbilder physischer Systeme erstellen. Diese digitalen Modelle ermöglichen die Simulation des Materialverhaltens unter unterschiedlichen Belastungs- und Umgebungsbedingungen. Auf dieser Grundlage lassen sich präzise Vorhersagen über die Lebensdauer und das Ausfallverhalten einzelner Komponenten innerhalb der Prozesskette treffen.

DOI 10.5162/iCCC2026/P16

Bild 6 Unsupervised Learning Quelle: Gantner Instruments GmbH

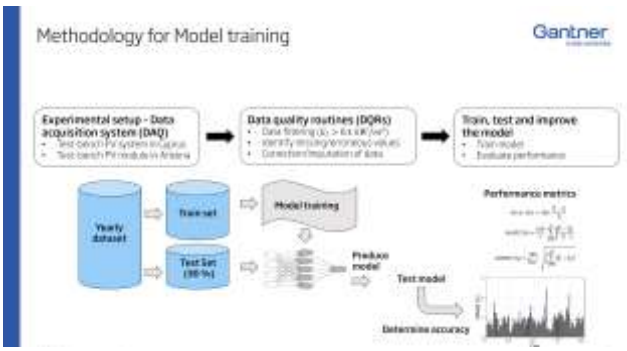


Bild 7 Modeltraining Quelle: Gantner Instruments GmbH

Die Integration von KI in diese Simulationsprozesse verbessert nicht nur die Genauigkeit der Prognosen, sondern reduziert auch den Bedarf an kostenintensiven physischen Tests. Darüber hinaus unterstützt sie die Entwicklung neuer Materialien und Fertigungstechnologien, indem sie komplexe Zusammenhänge zwischen Prozessparametern und Materialeigenschaften sichtbar macht.

Langfristig trägt die KI-gestützte Forschung dazu bei, die Innovationszyklen in der Halbleiterindustrie zu verkürzen und die Produktqualität zu steigern. Dies ist ein entscheidender Faktor, um den wachsenden Anforderungen an Miniaturisierung, Energieeffizienz und Zuverlässigkeit gerecht zu werden.

3 Diskussion und Ausblick

Die Ergebnisse der bisherigen Untersuchungen verdeutlichen, dass die Integration von Künstlicher Intelligenz (KI) die Prozesssicherheit und Effizienz in der Wafer-Produktion signifikant verbessert. KI-basierte Systeme ermöglichen nicht nur die Echtzeit-Analyse komplexer Datenströme, sondern auch die dynamische Anpassung von Prozessparametern, wodurch Produktionsfehler reduziert und die Ausbeute (Yield) gesteigert werden.

Zukünftige Entwicklungen wie **Edge-KI** und **digitale Zwillinge** werden die Automatisierung weiter vorantreiben. Edge-KI erlaubt die dezentrale Datenverarbeitung direkt am Messgerät, wodurch Latenzzeiten minimiert und die Reaktionsfähigkeit in kritischen Prozessen erhöht wird. Die Edge-KI benötigt Training, d.h. definierte Modelle werden von der Cloud auf das Edge Device gespeilt und arbeiten dort. Digitale Zwillinge schaffen virtuelle Abbilder physischer Systeme, die für Simulationen, prädiktive Analysen und die Optimierung komplexer Fertigungsabläufe genutzt werden können.

Ein besonderer Fokus liegt auf der **Nachhaltigkeit**, da optimierte Prozesse den Energieverbrauch und den Materialeinsatz erheblich reduzieren. KI-gestützte Systeme tragen somit nicht nur zur Effizienzsteigerung bei, sondern leisten auch einen wichtigen Beitrag zur Ressourcenschonung und zur Erfüllung regulatorischer Anforderungen im Hinblick auf den CO_2-Fußabdruck.

Die Zukunft der Messtechnik liegt in der engen Verzahnung von Hardware, Software und KI. Gantner Instruments nimmt in diesem Bereich eine Vorreiterrolle ein, indem Lösungen bereitgestellt werden, die über die reine Datenerfassung hinausgehen und ein tiefes Verständnis des gesamten Produktionsprozesses ermöglichen. Diese Technologien bilden die Grundlage für eine intelligente, adaptive und nachhaltige Industrieproduktion und ebnen den Weg für die nächste Generation der Halbleiterfertigung.

3.1 Zukunftstrends

Die zukünftige Entwicklung der Messtechnik in der Halbleiterindustrie wird maßgeblich durch drei zentrale Trends geprägt: **Edge-KI**, **digitale Zwillinge** und **Nachhaltigkeit**.

Edge-KI ermöglicht die Verarbeitung von Daten direkt am Messgerät, ohne dass diese zunächst an zentrale Server übertragen werden müssen. Dieser Ansatz reduziert Latenzzeiten erheblich und verbessert die Echtzeitfähigkeit der Systeme. Insbesondere in hochsensiblen Prozessen wie dem Schneiden von Siliziumwafern ist die sofortige Reaktion auf Anomalien entscheidend, um Qualitätsverluste und Produktionsunterbrechungen zu vermeiden. Edge-KI trägt somit zur Stabilität und Effizienz der Fertigung bei und reduziert die Abhängigkeit von Cloud-Infrastrukturen.

Digitale Zwillinge stellen virtuelle Abbilder physischer Systeme dar, die für Simulationen, Prozessoptimierungen und prädiktive Analysen genutzt werden können. Sie ermöglichen eine präzise Vorhersage des Systemverhaltens unter variierenden Bedingungen und unterstützen die Entwicklung adaptiver Steuerungsstrategien. In der Wafer-

Produktion können digitale Zwillinge beispielsweise genutzt werden, um die Auswirkungen von Temperatur- oder Vibrationsänderungen auf die Schnittqualität zu simulieren und die Prozessparameter entsprechend anzupassen.

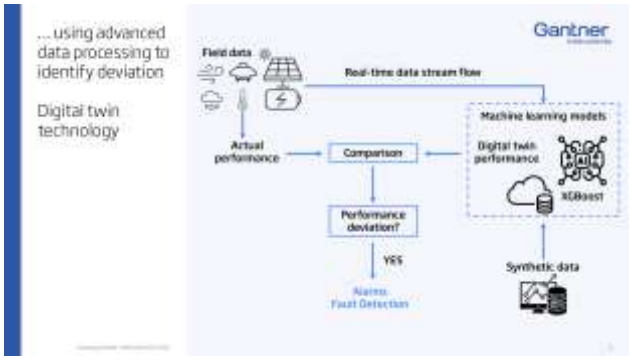


Bild 8 Digital Twin Technology Quelle: Gantner Instruments GmbH

Nachhaltigkeit wird durch den Einsatz KI-gestützter Systeme gefördert, da diese helfen, Ressourcen effizienter zu nutzen und den CO_2-Fußabdruck zu verringern. Optimierte Abläufe, vorausschauende Wartung und die Reduktion von Ausschuss tragen dazu bei, den Energieverbrauch und den Materialeinsatz zu minimieren. Dies ist nicht nur aus ökologischer Sicht relevant, sondern auch ein entscheidender Faktor für die Wettbewerbsfähigkeit in einer zunehmend regulierten Industrie.

Zusammenfassend lässt sich feststellen, dass diese Trends die Grundlage für eine intelligente, vernetzte und ressourcenschonende Produktion bilden. Sie eröffnen neue Perspektiven für die Automatisierung und die digitale Transformation der Halbleiterindustrie.

3.2 Fazit

Die fortschreitende Digitalisierung industrieller Prozesse erfordert eine enge Verzahnung von Messtechnik und datengetriebenen Analyseverfahren. Gantner Instruments positioniert sich in diesem Kontext als technologischer Vorreiter, indem jahrzehntelange Expertise in der hochpräzisen Datenerfassung mit innovativen Ansätzen der Künstlichen Intelligenz (KI) kombiniert wird. Die Produktplattformen **Q.series X**, **GI.cloud** und **Q.core** sind nicht lediglich Instrumente zur Messdatenerfassung, sondern integrale Bestandteile einer ganzheitlichen Architektur für intelligente Prozesssteuerung und datenbasierte Optimierung.

Die Integration von KI in Mess- und Prüfsysteme ermöglicht eine Transformation von rein reaktiven zu proaktiven und adaptiven Fertigungsprozessen. Durch die Echtzeit-Analyse großer Datenmengen können Anomalien frühzeitig erkannt, Wartungsmaßnahmen vorausschauend geplant und Prozessparameter dynamisch angepasst werden. Dies führt zu einer signifikanten Steigerung der Effizienz, einer Reduktion von Ausschuss und einer Erhöhung der Ausbeute (Yield), insbesondere in hochsensiblen Branchen wie der Halbleiterindustrie.

Darüber hinaus eröffnet die Kombination aus Messtechnik und KI neue Perspektiven für die Entwicklung digitaler Zwillinge, die als virtuelle Abbilder physischer Systeme dienen und für Simulationen sowie prädiktive Optimierungen genutzt werden können. Diese Ansätze tragen nicht nur zur Verbesserung der Prozesssicherheit bei, sondern leisten auch einen entscheidenden Beitrag zur Nachhaltigkeit, indem sie den Energieverbrauch und den Materialeinsatz minimieren.

Insgesamt sind die Lösungen von Gantner Instruments als Schlüsseltechnologien für die digitale Transformation zu betrachten. Sie schaffen die Grundlage für eine intelligente, vernetzte und ressourceneffiziente Industrieproduktion und ebnen den Weg für zukünftige Innovationen in den Bereichen Industrie, Energie und Forschung.

4 Literatur / Informationsquellen

[1] Gantner Instruments GmbH
[2] https://en.wikipedia.org/wiki/Edge_computing#Applications
[3] https://en.wikipedia.org/wiki/Deep_learning
[4] Patrick Schnabel: Elektrotechnik Fibel
[5] Europa Lehrmittel: Fachkunde Eletrotechnik
[6] Microsoft Copilot: Was ist Predictive Maintenance

DOI 10.5162/iCCC2026/P17

Fire risk assessment in open-cast mining areas using advanced sensing and signal processing at the edge of autonomous drones

Sharda Devi[1], Ireneusz Jablonski[1,2]

1) Brandenburg University of Technology (BTU), Cottbus-Senftenberg, Brandenburg
ShardaDevi@b-tu.de
2) Fraunhofer Institute of Photonics Microsystem (IPMS), Dresden, Saxony, Germany

Abstract

Recent advancements in drone technologies have increased their applicability in image-based fire detection. This paper presents image-based fire classification using convolution-based autoencoders (CAEs) and extended fusion models such as early fusion, feature-based fusion, and decision-based fusion. This study provides an overview of using fusion algorithms to classify fire using the FLAME3 dataset and a comparative study of the accuracy of the thermal dataset and the RGB dataset in the classification of fire. The results indicate that models trained on the thermal dataset could discriminate between *Fire* and *No Fire* instances better than models trained on the RGB dataset. Moreover, the CAE-based early fusion model achieved an *accuracy* around 95%, which is more as compared to feature level (92%) and decision level (91%). It can be concluded that fusion approaches are credible for fire classification and have the potential to localize fire-prone areas. Future studies aim to achieve the classification and localization of fire risks in mined-out areas in open-cast mining using different sensors on drone platforms. As a guide for future work, the focus would first be to reduce the complexity and size of AI-based models using pruning, quantization, and knowledge distillation so that they can be used on drone platforms to run real-time inference. As a collaborative edge-cloud-based approach, edge AI (artificial intelligence) servers are proposed to be to run these lightweight models onboard, and computationally expensive tasks and processing can be shifted to cloud platforms.

1 Introduction

Drones are nowadays extensively used in monitoring critical and complex infrastructures, precision agriculture, search and rescue operations, etc. Mining areas are also typically vast and structurally complex, consisting of uneven terrain, dynamic operational zones, and diverse environmental conditions. Because of this heterogeneity, a single sensor cannot provide comprehensive coverage of fire scenes in these sites. Therefore, to assess the fire scenario in large open-cast mines, multiple sensors that can detect physical quantities such as environmental variations, smoke presence, and moisture levels [1] can be used. State-of-the-art fire detection methods commonly combine RGB and thermal (IR) cameras on drones, showing that multimodal sensing improves detection reliability and lowers false-alarm rates. Inspired by these findings, this study explores the effectiveness of applying a similar multimodal strategy for fire detection in mining environments. A CAE based architecture [2] is employed to perform a comparative analysis between RGB and IR image representations. There is extensive literature available that employs the use of AI approaches on UAV platforms to detect incidents of fire [3]. Running AI models onboard using GPUs consumes more power [4], and more computation reduces battery life [5], which makes UAVs resource constrained. This research study proposed integrating edge and cloud computing for drone platforms to deal with the large amounts of data coming from multiple sensors and real-time detection.

2 State of the art

2.1 Multimodal Data Fusion

In open-cast mining, spontaneous combustion can occur, especially in lignite regions, where fine powder is moved by wind and causes a chain reaction to start continuous ignition [6]. Additionally, after finishing mining in certain regions, machines in the open-cast mining area are scraped with burners and dynamite; the embryos fall on the lignite surface. These embryos are flammable after some hours, which can lead to the fire in that region. Timely detection of fire can save lives of mining workers, minimize environmental damage, and facilitate operational continuity. Hence, continuous monitoring is required to inspect the fire regions.

Drones can be equipped with multiple sensors apart from camera sensors, such as gas sensors [2], temperature sensors, RADAR [7], LiDAR [7], etc. The data from these sensors can be fused using machine learning and deep learning algorithms to gain a complete understanding of fire scenes. This is called multi-sensor data fusion [6]. Different AI methodologies, like neural network-based models, transformer-based vision models, and generative models, etc., can be used, which are tailored for efficiency and robustness [7]. This multisensory approach enhances the accuracy of fire probability models and reduces false alarms. This system can be extended to real-time alerts [8] for the fire department to assess the areas and deploy an emergency response.

2.1.1 Encoder-decoder based fusion

Convolutional based autoencoder (CAE) is a variant of CNNs [2]. It is a special type of neural network that compresses data in compact form and reconstructs it to closely match the original input (*Figure 1*). The reconstruction error can be calculated when the model is tested on the same patterns or images that are not used for training. Based on the reconstruction error, classification tasks can be accomplished.

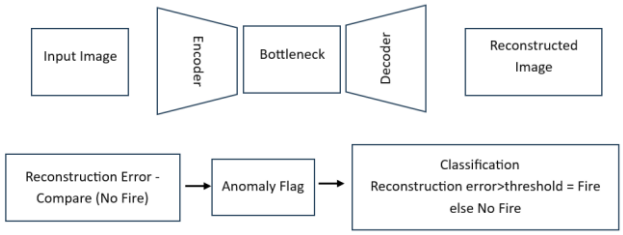


Figure 1 CAE based autoencoder algorithm to discriminate between *Fire* and *No Fire*

The fusion approaches used to analyze multimodality include early fusion, feature-level fusion, and late fusion (*Figure 2*) [2]. In the early fusion approach, different modalities are combined before being sent to the encoder, allowing the model to extract the high-level features from all modalities. These high-level features are used by the decoder to reconstruct the original inputs. Feature-level fusion is performed after features are independently extracted from each modality using separate encoders; the encoded representations are then fused and then passed to the decoder or classifier for decision-making. In the late fusion approach, separate CAE based classifiers are trained for each modality, and their outputs are combined, such as by majority voting or weighted average, to produce the final classification result.

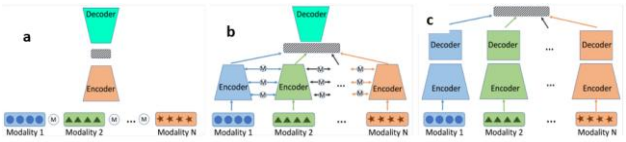


Figure 2 CAE based fusion approaches - a) Early - b) Feature level (mid-) c) Decision level (late-) fusion

2.2 Cloud Computing vs Edge Computing

Data collection and processing is traditionally performed on cloud servers (*Figure 3*). The cloud supports extensive data storage, analysis, and remote monitoring [8]. Edge AI, on the other hand, takes artificial intelligence on edge devices. It extends the concept of cloud computing by bringing all data collection and processing closer to the end user at the network edge. As a result, data transmission distances are shorter, leading to reduced latency and minimal transmission losses compared to traditional cloud-based processing.

A modular system architecture that combines cloud and edge computing can offer benefits of computing strategies [7, 8]. The cloud can handle data storage, large-scale processing, and model updates while real-time inference is performed on the edge. This hybrid approach improves scalability and availability in the fire risk assessment system. The cloud side [8] provides centralized data management and algorithm synchronization, whereas the edge side [1] enables low-latency local detection. This design delivers an efficient, robust, and responsive solution capable of supporting data from different modalities for intelligent fire monitoring.

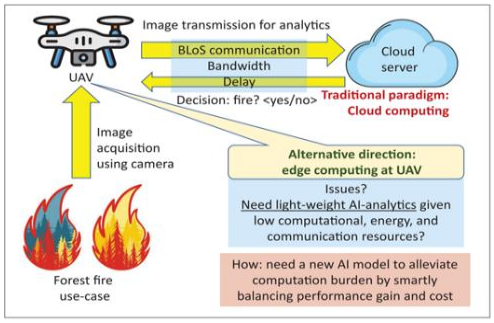


Figure 3 Forest fire use case - there are delays in fire detection because of bandwidth delay in processing on cloud servers and this delay is reduced by bringing processing drone platform with edge computing

3 Materials and Methods

CAE based autoencoders are implemented on the FLAME3 dataset (*Table 1*), which is based on the reconstruction of input images [9]. The detection mechanism is exploited to flag fire images as anomalous images. In the first step, the model is trained on RGB and thermal images separately (*Figure 4*).

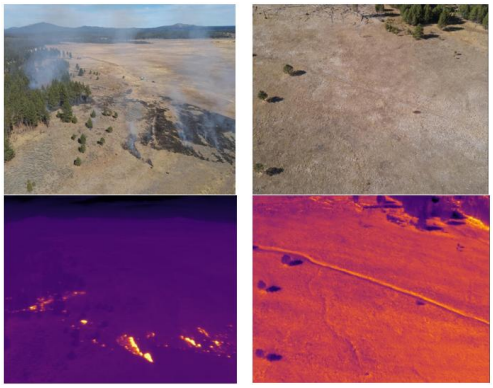


Figure 4 FLAME3 dataset - Thermal and RGB images of *Fire* and *No Fire* instances

In the second step, the fusion approaches used to analyze multimodality include early fusion, feature-level fusion, and late fusion. The input images were resized to 64×64 pixels, and the model was trained using a batch size of 32 for 10 epochs. The encoder is compressed to 32- dimensional latent space, and the network was optimized with a learning rate of 0.001.

The confusion matrix (*Figure 5a*) evaluates the performance of a classification model. It compares the model's predictions against the actual ground truth, illustrating where the model is succeeding or failing. This detailed breakdown is essential for computing other performance

DOI 10.5162/iCCC2026/P17

metrics such as *accuracy, precision, recall* and *F1 score*, which are derived from its core components (*Figure 5b*).

a

Confusion Matrix		
True: No Fire	True Positive (TP)	False Positive (FP)
True: Fire	False Negative (FN)	True Negative (TN)
	No Fire	Fire

Predicted

b

$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$	$Recall = \frac{TP}{TP+FN}$
$Precision = \frac{TP}{TP+FP}$	$F1 = 2 \times \frac{Precision \times Recall}{Precision+Recall}$

Figure 5 a) Confusion matrix b) Performance matrices derived from confusion matrix to evaluate the correctness of models

4 Results

When trained individually, the thermal-only autoencoder was more accurate than the RGB-only model. As can be seen in *Figure 6*, there are more instances where fire is misclassified as no fire when the model is tested on the RGB dataset, but misclassifications decrease in the case of the thermal dataset. The underlying reason is that the convolution autoencoder decodes the hotspots from thermal images, which can be directly correlated with the fire. It is naturally a good discriminative signal. The fire scenes from RGB images can be ambiguous because of visible light, smoke, reflections, or bright sunlight. It can sometimes look like a fire.

Dataset	Train (No Fire)	Validation (No Fire)	Test (Fire)
FLAME3	66	50	622

Table 1 FLAME3 dataset

The dataset contains 622 fire images and 116 non-fire images across thermal and RGB modalities. Of the non-fire images, 66 are used for training and 50 for estimating reconstruction errors. The model is then evaluated on 622 fire samples. Because the dataset is highly imbalanced, accuracy is not a reliable metric. Instead, *Precision, Recall,* and *F1-score* are computed for both the *Fire* and *No-Fire* classes, and their average is reported as the overall performance (*Table 2a*).

By taking the average of *Recall* from both classes, balanced accuracy can be calculated.

$$Balanced\ accuracy = \frac{Recall\ (Fire) + Recall\ (No\ Fire)}{2}$$

In the early fusion approach, RGB and thermal images are concatenated at the input level. It allows the model to learn joint representations from both the modalities. It results in better correlation and improved discrimination between fire and no-fire regions (*Figure 7*). This is evident from *Table 2b*; the early fusion convolutional autoencoder achieved higher *accuracy* and *precision* of 95% and 99% respectively as compared to mid- and decision level fusion having *accuracy* 92% and 91% respectively, suggesting that combining inputs at the earliest stage allowed the model to learn a more robust joint representation of normal versus fire scenes.

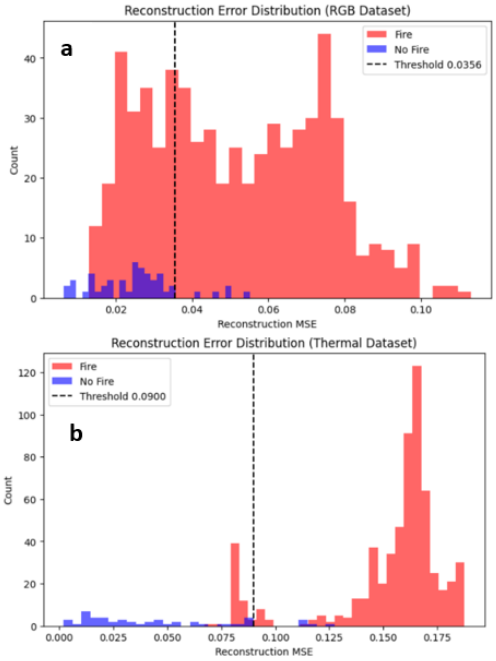


Figure 6 Reconstruction error distribution a) RGB dataset b) Thermal dataset

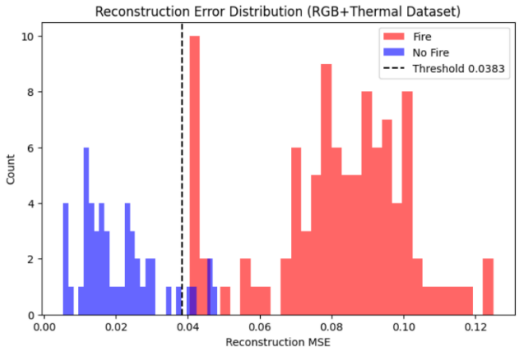


Figure 7 Reconstruction error distribution - Early fusion

a

	Metric	*Precision*	*Recall*	*F1 Score*
Early Feature level	Fire	0.99	0.96	0.98
	No Fire	0.66	0.94	0.78
Decision level	Fire	0.99	0.92	0.96
	No Fire	0.49	0.98	0.65
	Fire	0.98	0.92	0.95
	No Fire	0.46	0.82	0.59

b

Fusion	*Accuracy*	*Precision*	*Recall*	*F1-score*
Early	0.959	0.995	0.961	0.977
Feature level (Mid)	0.921	0.998	0.916	0.955
Decision level (Late)	0.915	0.984	0.922	0.952

Table 2 a) Performance of fusion models on *Fire* and *No Fire* instances b) Combined performance matrices

5 Discussion and outlook

Preliminary results of performed research indicate that the robustness of the early fusion-based CAE as compared to feature-level and decision-level fusion. The next logical step is deploying CAE fusion models on UAV platforms. It enables real-time, localized fire risk assessment with low latency and minimal dependency on communication channels. According to theoretical analysis, early fusion has the lowest latency and energy consumption [2], making it ideal for real time onboard inference. However, it could lose modality-specific details in contrast to feature level and decision level fusion. But feature level and decision level based fusion algorithms can add more computation cost, making them relatively slower. To deploy these models on resource-constrained UAVs, these models should be lightweight. A typical strategy to make the model lightweight in quantization (INT8), pruning, and knowledge distillation [5].

Recent research studies [4, 5, 8] have shown that it is now possible to use edge and hybrid AI architectures to detect fires in real time on small, low-power devices. High-performance edge devices like NVIDIA Jetson Xavier NX or Orin Nano may achieve real-time onboard inference at 60 frames per second with an end-to-end latency of ~12-16 ms per frame. This includes data collecting, preprocessing, inference, and postprocessing. Memory footprint can vary from 200 to 350 MB for model weights and activation, which can fit comfortably with Jetson's 2-4 GB available GPU/CPU memory. Early fusion is expected to be optimal for strict latency considerations, while mid-fusion models may offer slightly higher accuracy at a modest latency cost.

6 Conclusion

CAE-based fusion approaches highlight that early fusion of RGB and thermal modalities enables the network to learn richer joint feature representations and enhances the robustness of fire detection, while thermal imagery alone remains a strong and stable modality for detecting fire in varying lighting and environmental conditions. Since the main aim is to deploy this multi-sensor system on a drone platform, it requires more computation and space, which is more challenging to have well-developed drones. To ensure that models can fit within the limited memory on edge devices, model compression techniques such as pruning, quantization, and low-rank factorization can be used. These techniques minimize the model size while maintaining acceptable levels of accuracy and performance. In the following, dedicated sensors can be integrated onboard UAVs for real-time data acquisition and inference onboard using distributed cloud and edge servers.

7 Literature

[1] M. M. Fouda, S. Sakib, Z. M. Fadlullah, N. Nasser, and M. Guizani, "A Lightweight Hierarchical AI Model for UAV-Enabled Edge Computing with Forest-Fire Detection Use-Case," *IEEE Network*, vol. 36, no. 6, pp. 38–45, 2022, doi: 10.1109/MNET.003.2100325.

[2] F. Zhao, C. Zhang, and B. Geng, "Deep Multimodal Data Fusion," *ACM Comput. Surv.*, vol. 56, no. 9, pp. 1–36, 2024, doi: 10.1145/3649447.

[3] A. Abdusalomov *et al.,* "Drone-Based Wildfire Detection with Multi-Sensor Integration," *Remote Sensing*, vol. 16, no. 24, p. 4651, 2024, doi: 10.3390/rs16244651.

[4] U. U. Deshpande *et al.,* "Real-time fire and smoke detection system for diverse indoor and outdoor industrial environmental conditions using a vision-based transfer learning approach," *Front. Comput. Sci.*, vol. 7, 2025, doi: 10.3389/fcomp.2025.1636758.

[5] S. Mahmoudi, M. Gloesener, M. Benkedadra, and J.-S. Lerat, "Edge AI System for Real-Time and Explainable Forest Fire Detection Using Compressed Deep Learning Models," in *Proceedings of the 20th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*, Porto, Portugal, 2025, pp. 847–854.

[6] X. He, X. Yang, Z. Luo, and T. Guan, "Application of unmanned aerial vehicle (UAV) thermal infrared remote sensing to identify coal fires in the Huojitu coal mine in Shenmu city, China," *Scientific reports*, vol. 10, no. 1, p. 13895, 2020, doi: 10.1038/s41598-020-70964-5.

[7] M. Yang, S. Qian, and X. Wu, "Real-time fire and smoke detection with transfer learning based on cloud-edge collaborative architecture," *IET Image Processing*, vol. 18, no. 12, pp. 3716–3728, 2024, doi: 10.1049/ipr2.13187.

[8] X. Yang, Y. Li, and Q. Chen, "Automated image-based fire detection and alarm system using edge computing and cloud-based platform," *Internet of Things*, vol. 28, p. 101402, 2024, doi: 10.1016/j.iot.2024.101402.

[9] B. Hopkins *et al.,* "FLAME 3 Dataset: Unleashing the Power of Radiometric Thermal UAV Imagery for Wildfire Management," Dec. 2024. [Online]. Available: http://arxiv.org/pdf/2412.02831v1

DOI 10.5162/iCCC2026/P18

Early Prediction of H_2 Concentration and Sensor Response Using a Multi-Layer Perceptron (MLP)

Raduan Sarif and Carlo Tiebe
Bundesanstalt für Materialforschung und -prüfung (BAM)
Berlin, Germany
Email: raduan.sarif@bam.de; carlo.tiebe@bam.de

Christian Herglotz
Brandenburgische Technische Universität Cottbus-Senftenberg
Cottbus, Germany
Email: christian.herglotz@b-tu.de

Abstract

The hydrogen (H_2) economy is expanding rapidly, promising to be a clean energy solution; however, challenges related to flammability and gas leakage during storage and transportation remain critical. Early monitoring of rising H_2 concentration is crucial for industrial safety and leakage detection; however, in practice, sensor outputs can be subject to drift from the real concentration, resulting in a calibration or adjustment being necessary for an appropriate alarm. To address this limitation, this work presents a Multi-Layer Perceptron (MLP) model that predicts the H_2 concentration and the entire sensor response, including the stable value, using only a 3-second early-time window of the sensor output signal. The proposed approach enables prediction of the final concentration 62% to 92% faster than the traditional t_{90} response time across different H_2 concentration levels. These results demonstrate the potential of fast, model-based prediction to enhance both the speed and reliability of H_2 monitoring systems, thereby supporting the safer deployment of H_2 technologies in real-world applications.

Die Wasserstoffwirtschaft (H_2) gilt als wichtiger Baustein für die Transformation hin zu einer kohlenstofffreien Energieversorgung. Dennoch gibt es Anforderungen für einen sicheren Umgang mit dem Energieträger Wasserstoff zur berücksichtigen. Entlang der Wertschöpfungskette stellen Leckagen ein Sicherheitsdefizit dar. Eine frühzeitige Erkennung ansteigender H_2-Konzentrationen mit Gasdetektoren/Gassensorsystemen, insbesondere in abgeschlossenen Räumen oder zum Freimessen von Komponenten, ist daher von hoher Bedeutung. Diese Arbeit beschreibt die Anwendung des datenanalytischen Modells „Multi-Layer-Perceptron (MLP)“ zur Nutzung eines 3 Sekunden langen Intervalls des Sensorsignals (Zeitreihensignals), um sowohl die H_2-Konzentration als auch den vollständigen Signalverlauf vorherzusagen. Der Ansatz ermöglicht eine Abschätzung der finalen Konzentration um 62% bis 92% schneller als alternative Verfahren zur Beschreibung des transienten Sensorsignalverhaltens. Die Ergebnisse zeigen, dass eine modellbasierte Vorhersage die Geschwindigkeit und Zuverlässigkeit von H_2-Überwachungssystemen erhöht. Die Verwendung dieses Modells in realen Anwendungen trägt somit zur Sicherheit von Wasserstoffgassystemen bei.

1 Introduction

Hydrogen monitoring is essential in storage systems, pipelines, electrolyzers, fuel cells, and refueling stations, where escaping H_2 can accumulate to explosive levels. A reliable monitoring system capable of detecting rising concentrations is therefore critical. Because hydrogen is highly flammable, issuing an alarm before it reaches the ignition threshold—the lower explosion limit (LEL) of 4 Vol% in air [1]—is vital. If the H_2 concentration can be predicted at an early stage, timely safety actions can be initiated. The predicted steady-state sensor response further supports rapid decision-making, thereby reducing the risk of severe incidents.

Recent developments in neural networks have significantly advanced gas sensor technologies, enabling qualitative and quantitative gas identification [2]. In addition to improving accuracy, neural network algorithms have demonstrated the ability to rapidly identify appropriate gases and predict responses by analyzing the transient signals of the sensors [3, 4, 5, 6]. Although these studies show the potential of data-driven approaches, only a few have addressed the prediction of H_2 concentration using a small time window from the early response of the sensor and real H_2 flow, which enables predicting critical H_2 concentrations before they occur [7, 8, 9]. By using the early-time window values of the sensor signal, this study introduces a novel application of the MLP model for three key purposes: (1) predicting the entire sensor response, (2) estimating the stable sensor value, and (3) estimating the real H_2 concentration. This

DOI 10.5162/iCCC2026/P18

concept builds on our previous work [10] [11] [12], using a first-order (FO) exponential model and MLP to predict a stable sensor response from a small time window. This study uses a 3 s early-time window to predict both the sensor response and the H_2 concentration, which falls within the 2 s to 5 s range required for various H_2 safety-critical applications recommended in [13]. Additionally, this work introduces new sensor response datasets, enabling the proposed model to achieve up to 85% faster prediction performance on real sensor time-series data compared to previous methods.

The remaining sections of this paper are structured as follows: Section II describes the data acquisition process; Section III outlines the MLP methodology; Section IV presents the prediction and evaluation results; Section V discusses the model's prediction accuracy; and Section VI concludes the paper.

2 Data Acquisition

For data acquisition in this study, we used the XEN-5320 H_2 thermal conductivity sensor type, whose datasheet specifies a response time of less than 1 s and a t_{90} value of under 5 s [14]. The sensor was mounted on the XEN-85030 gas manifold, which has a volume of (60 mm × 80 mm x 18 mm) and is equipped with standard 5-mm inlet and outlet ports [15]. In accordance with the DIN EN ISO 6145-6 standard [16], H_2 was mixed with air to generate concentrations from 0.1 to 2.0 Vol% at total flow rates of 1000, 1500, 2000, and 3000 mL/min. Each H_2 concentration level duration was maintained for 1 hour, and the sensor's sampling frequency was 1 Hz.

In Fig. 1, each block displays the sensor-measured time series (green) (Vol%) alongside the corresponding H_2 concentration (orange) (Vol%) collected from the mass-flow controller (MFC). In the following figure, the x-axis represents time in hours, while the y-axis shows the sensor response and the corresponding H_2 concentration in Vol%. For total flow rates of 1000, 1500, and 2000 mL/min, 20 experimental time series were recorded. For the 3000 mL/min, only eight time series were obtained due to an MFC malfunction.

Figure 2 presents an example where the given H_2 concentration is 1.0 Vol% at a total flow of 1000 mL/min. In this case, the sensor response (Vol%) (green) reaches approximately 1.45 Vol% for a real concentration of 1.0 Vol% (orange), demonstrating a noticeable bias relative to the actual H_2 level. To address this issue, the proposed MLP model is designed not only to predict the stable sensor response but also to estimate the true underlying H_2 concentration.

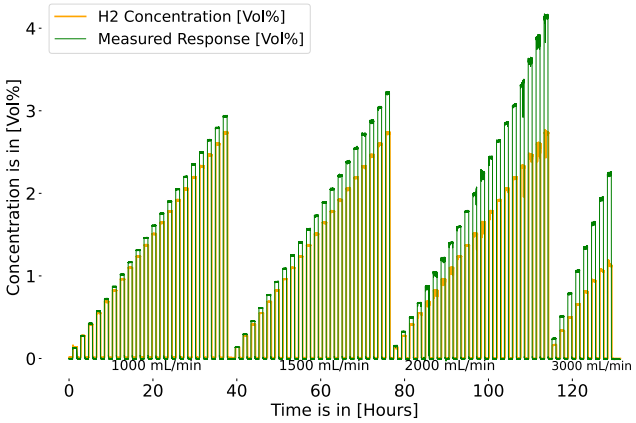


Figure 1: Measured sensor response and corresponding given H_2 concentration (for 0.1 Vol% 2.0 Vol%) at a total flow rate of 1000,1500,2000,3000 mL/min.

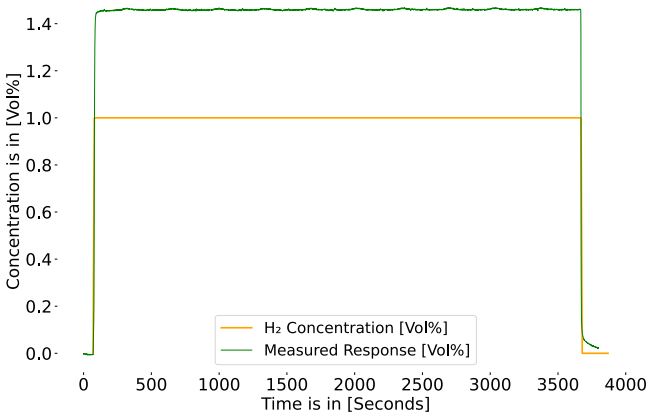


Figure 2: Measured sensor response and corresponding given H_2 concentration (for 1.0 Vol%) at a total flow rate of 1000 mL/min.

3 Methodology of the MLP

In this study, we employ an MLP model [17], whose input layer receives the raw sensor values and forwards them directly to the hidden layers. Each neuron represents one feature of the input data, so the number of input neurons matches the dimensionality of the input vector. The MLP input I_p consists of sensor values extracted from a predefined small-time window, as described in Eq. (1). These input features-i.e., the small time-window samples $S_{\mathrm{w}}(t)$-begin at the time index where the sensor output exceeds the threshold S_{th} and end at the time index t_{w}. The MLP output layer predicts the target quantities defined in Eq. (2), namely sensor time series response up to 600s, the stable sensor value $\hat{S}_{\mathrm{stable}}$ and the corresponding H_2 concentration Y_{H_2}.

$$\boldsymbol{I}_p = [S_w(t_1), S_w(t_2), \ldots, S_w(t_w)] \tag{1}$$

DOI 10.5162/iCCC2026/P18

$$\mathbf{O}_p = \{\hat{S}(t),\ \hat{S}_{\text{stable}},\ Y_{\text{H}}\} \tag{2}$$

In the MLP input and output layers are connected through hidden layers, as illustrated in Fig. 3. Each input to a hidden layer neuron is weighted by a corresponding weight W and added to a bias b. The weights determine the strength of the connections between neurons, while the bias shifts the activation function, allowing the model to adjust the activation threshold. The MLP learns weights and biases during the training process to adaptively fit the training data and capture complex patterns of the datasets.

In Eq. (3), each hidden neuron $h_{j,k}$ computes a weighted sum of the inputs received from the previous layer. Here, k denotes the time index t_w of the input features, and j denotes the index of the neurons in the hidden layer. Here, $I_{p(k)}$ represents the input features at time index k. The resulting weighted sum is then passed through an activation function, typically the Rectified Linear Unit (ReLU), denoted as $f(h_{j,k})$ and defined in Eq. (4). In the subsequent hidden layers, each neuron performs the same computation, but using the outputs of the previous layer instead of the raw sensor inputs. The proposed model uses two hidden layers, each containing $j = 162$ neurons.

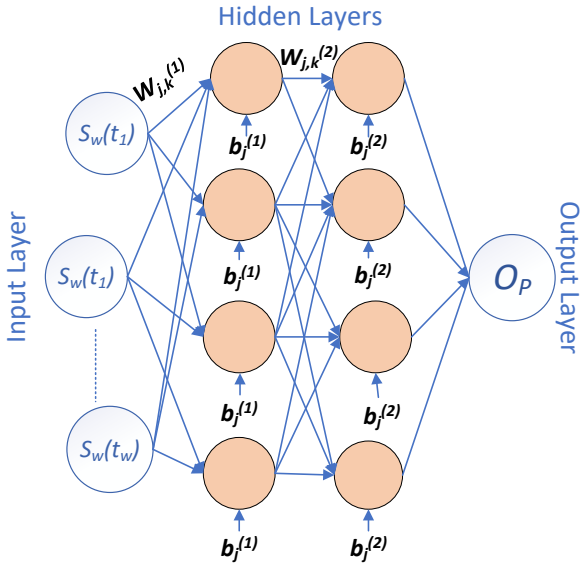


Figure 3: MLP architecture with an input layer, two hidden layers with separate weights for each neuron ($W_{j,k}^{(1)}, W_{j,k}^{(2)}$) and biases ($b_j^{(1)}, b_j^{(2)}$), and an output layer ($O_p$) for stable sensor response prediction [12].

$$h_{j,k} = \sum_{1}^{t_w} I_{p(k)} \cdot W_{j,k} + b_j \tag{3}$$

$$f(h_{j,k}) = \begin{cases} h_{j,k}, & \text{if } h_{j,k} > 0 \\ 0, & \text{if } h_{j,k} \leq 0 \end{cases} \tag{4}$$

Before training the MLP, a Python-based algorithm was developed to automatically detect segments of the sensor time series $S_{\text{r}}(t)$ and the corresponding given H_2 concentration X_{H_2} reported by MFC. The algorithm identifies rising edges where the signals exceed predefined thresholds: $S_{\text{th}} > 0.02$ Vol% for $S_{\text{r}}(t)$ and $X_{H_2}(t) > 0.1$ Vol% for the concentration signal. For each detected segment, a 3600-second sample of both $S_{\text{r}}(t)$ and X_{H_2} is extracted for further processing. These detected segments of $S_{\text{r}}(t)$ and $X_{H_2}(t)$ are then used to construct the training dataset for the MLP. From each segment, the model extracts a short early-time window, the stable sensor value, and the corresponding H_2 concentration.

During testing, the model receives an early-time window from seven randomly selected sensor response time series, representing 10% of the total dataset. This window is scaled and provided as input to generate the predicted sensor response. The predicted curve starts from the last point of the early-time window and extends up to 600 seconds. In the inference stage, the model receives an input vector I_p derived from the early-window features $S_{\text{w}}(t)$ and produces the output vector O_p.

We evaluate the MLP prediction accuracy using the relative fitting error ε defined in Eq. (5), computed between the real sensor response $S_r(t)$ and the model prediction $\hat{S}(t)$. The error is calculated over discrete time indices t_i with a total of N samples. As well as the stable-value prediction error ε_s, shown in Eq. (6) is the relative absolute difference between the true stable value S and its prediction $\hat{S}$, expressed as a percentage.

Similarly, Eq. (7) defines the H_2 concentration prediction error, obtained by comparing the given concentration X_{H_2} with the model-predicted concentration Y_{H_2}. Finally, the relative time savings η_{s} are computed using Eq. (8) by comparing the model's prediction time $\hat{t}$ with the sensor's t_{90} response time. The metric t_{90}, a widely used indicator of sensor response speed [18, 16], was determined in this study through graphical analysis. The total prediction time $\hat{t}$ consists of the early-time window duration t_w plus the MLP inference time. Since the inference time is only a few milliseconds, it was not included in the analysis of this study.

$$\varepsilon = \frac{1}{N} \sum_{i=1}^{N} \left(\frac{|S_r(t_{\text{i}}) - \hat{S}(t_{\text{i}})|}{S_r(t_{\text{i}})} \right) \times 100\% \tag{5}$$

DOI 10.5162/iCCC2026/P18

$$\varepsilon_s = \left| \frac{S - \hat{S}}{S} \right| \times 100\%. \tag{6}$$

$$\varepsilon_{H_2} = \left| \frac{X_{H_2} - Y_{H_2}}{X_{H_2}} \right| \times 100 \tag{7}$$

$$\eta_s = \left(\frac{t_{90} - \hat{t}}{t_{90}} \right) \times 100\% \tag{8}$$

4 Results Analysis

The model validation dataset was generated separately at H_2 concentrations of 0.3, 0.5, 0.7, 1.0, 1.5, and 2.0 Vol% under a total flow rate of 1000 mL/min, using the same experimental procedure. For prediction, the input early-time window $S_w(t)$ consisted of the early 3-second samples above the threshold $S_{\text{th}} > 0.02$ Vol% for each concentration. The MLP predicts the sensor response $\hat{S}(t)$ up to 600 s, including the stable value $\hat{S}$ and the estimated H_2 concentration Y_{H_2}. Figure 4 shows the 1 Vol% result, where the horizontal axis represents time in seconds, and the vertical axis represents the sensor output in Vol%. The blue curve is the 3 s early-time window $S_w(t)$, the green curve is the measured response stable value $S = 1.31$ Vol%, and the red curve is the predicted response $\hat{S}(t)$ (stable value 1.203 Vol%).

The predicted response $\hat{S}(t)$ shows some fluctuations and is slightly lower than the measured stable value. This is expected given the limited size of the training dataset and the simple MLP architecture. Even with these constraints, the model could learn the critical dynamic trend, and the remaining noise and lower estimation can be further improved by training on larger datasets or using a more complex network.

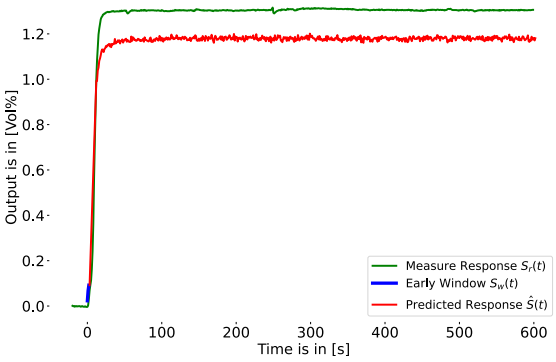


Figure 4: Measured and MLP-predicted sensor response for 1 Vol% H_2 for total flow of 1000 mL/min.

5 Discussion

Table 1 presents the overall MLP validation results for different H_2 concentrations. The real concentration is denoted by X_{H_2} and the predicted concentration by Y_{H_2} (Vol%). Similarly, the real stable sensor output is represented by S and its predicted value by $\hat{S}$ (Vol%), with the corresponding stable-value error ε_s defined in Eq. (6). The H_2 concentration error ε_{H_2} is computed using Eq. (7), and the model relative fitting error ε is calculated using Eq. (5). The relative time savings η_s are determined from Eq. (8), where the prediction time was fixed at $\hat{t} = 3$ s for all H_2 concentrations.

Table 1: Overall MLP Validation Summary for Different H_2 Concentrations

X_{H_2}	Y_{H_2}	ε_{H_2}	S	$\hat{S}$	ε_s	t_{90}	η_s	ε
0.3	0.25	19.00	0.42	0.35	17.27	40	92.68	19.84
0.5	0.49	3.00	0.72	0.58	19.44	31	90.32	15.01
0.7	0.83	18.57	1.01	1.00	1.38	17	82.35	17.14
1.0	1.10	10.30	1.31	1.20	16.18	12	89.00	18.19
1.5	1.60	6.13	2.18	1.94	11.24	10	70.00	17.10
2.0	2.30	18.25	2.92	2.64	9.69	8	62.50	18.09
Mean	–	**12.87**	–	–	**12.53**	–	**81.14**	**17.56**

The MLP time-series prediction demonstrates that the model can learn the rising trend of the H_2 sensor response using only a 3-second early-time window, achieving an accuracy range of 80% to 85% (mean error: 17.56%). The stable-value prediction further demonstrates that the model can reliably estimate the final steady-state level across all validated concentrations, with an accuracy of 83% to 97% (mean error: 12.53%). Overall, the predicted curves follow the general sensor dynamics, reach the stable values, and reproduce the gradual increase toward convergence. In the predicted response curve, some fluctuations are observed, primarily due to the limited size and variability of the real training dataset, as well as the simplicity of the MLP architecture.

In addition, the model also predicts the real H_2 concentration, achieving 80% to 97% accuracy (mean error: 12.87%). Although errors increase at different concentrations due to non-linearities in the MFC values and fewer training samples, the model still reduces the offset between the sensor output and the real H_2 concentration. This ability to estimate the real concentration from only a small early-time window is highly valuable for fast and reliable early-alarming systems. Above all, early-window prediction of the H_2 sensor response can save response time by 62% to 92% (mean time savings 81.14%), enabling much faster H_2 concentration detection compared to the conventional t_{90} time.

6 Conclusion

Overall, the primary objective of this study is to use only the early-time window values of the sensor response to predict the real H_2 concentration, as well as the entire transient sensor response and the final stable value, which are used together as the ground truth for evaluation. By predicting the sensor's entire response before it reaches full convergence, the model can save up to 90% of the time compared to the response time of conventional sensors. This approach enables a fast, reliable, and intelligent H_2 measurement system that enhances industrial safety. It supports future applications such as early leak detection, real-time H_2 concentration monitoring, and a digital twin for hydrogen refueling stations. Since the model estimates the real concentration before the sensor response reaches t_{90}, in any hazardous situation, an alarm system can be triggered earlier and more reliably. In future work, this approach will be validated using a wider range of H_2 concentrations and sensor responses, including the integration of multiple sensors from different sensing principles to improve robustness and generalization.

References

[1] M. Molnarne and V. Schroeder, "Hazardous properties of hydrogen and hydrogen containing fuel gases," *Process Safety and Environmental Protection*, vol. 130, pp. 1–5, 2019.

[2] B. Zong, S. Wu, Y. Yang, Q. Li, T. Tao, and S. Mao, "Smart gas sensors: recent developments and future prospective," *Nano-Micro Letters*, vol. 17, no. 1, p. 54, 2025.

[3] D. L. Osorio-Arrieta, J. L. Muñoz-Mata, G. Beltrán-Pérez, J. Castillo-Mixcóatl, C. O. Mendoza-Barrera, V. Altuzar-Aguilar, and S. Muñoz-Aguirre, "Reduction of the measurement time by the prediction of the steady-state response for quartz crystal microbalance gas sensors," *Sensors*, vol. 18, no. 8, p. 2475, 2018.

[4] R. R. Patil, R. K. Calay, M. Y. Mustafa, and S. Thakur, "Artificial intelligence-driven innovations in hydrogen safety," *Hydrogen*, vol. 5, no. 2, pp. 312–326, 2024.

[5] T. Hübert, J. Majewski, U. Banach, M. Detjens, and C. Tiebe, "Response time measurement of hydrogen sensors," *Hydrogen Knowledge Centre*, 2017.

[6] C. Shi, W. Pei, C. Jin, A. Alizadeh, and A. Ghanbari, "Prediction of the sno2-based sensor response for hydrogen detection by artificial intelligence techniques," *International Journal of Hydrogen Energy*, vol. 48, no. 52, pp. 19834–19845, 2023.

[7] Y. Shi, S. Ye, and Y. Zheng, "Rapid forecasting of hydrogen concentration based on a multilayer cnn-lstm network," *Measurement Science and Technology*, vol. 34, no. 6, p. 065101, 2023.

[8] V. Martvall, H. Klein Moberg, A. Theodoridis, D. Tomecek, P. Ekborg-Tanner, S. Nilsson, G. Volpe, P. Erhart, and C. Langhammer, "Accelerating plasmonic hydrogen sensors for inert gas environments by transformer-based deep learning," *ACS sensors*, 2025.

[9] G. Yang, D. Kong, X. He, X. Yu, and K. Jiang, "Prediction of hydrogen leakage location and intensity in hydrogen refueling stations based on deep learning," *International Journal of Hydrogen Energy*, vol. 68, pp. 209–220, 2024.

[10] R. Sarif, C. Tiebe, and C. Herglotz, "Early response prediction for h2 sensors," in *Proceedings of the 2025 IARIA Annual Congress on Frontiers in Science, Technology, Services, and Applications (IARIA Congress 2025)*, 2025.

[11] R. Sarif, C. Tiebe, and C. Herglotz, "Analysis of methods for predicting h_2 sensor responses," in *Proceedings of the Sensor Measurement Science International (SMSI) Conference*, (Nürnberg, Germany), pp. 169–170, AMA Science, May 2025.

[12] R. Sarif, C. Tiebe, and C. Herglotz, "Early H_2 sensor response prediction using a multi-layer perceptron (mlp)," in *Proceedings of the IEEE Sensors Conference*, (Vancouver, Canada), IEEE, 2025.

[13] T. Hubert, U. Banach, S. Machill, and P. Moretto, "Summary report for a hydrogen sensor workshop," tech. rep., European Commission, Joint Research Centre (JRC), 2017.

[14] Xensor Integration, *XEN-5320-ALU: Thermal Conductivity Gas Sensor*, 2020. Version 30 September 2020.

[15] Xensor Integration, *XEN-85030: Gas Manifold for XEN-5320 Sensor*, 2020. Version 30 September 2020.

[16] International Organization for Standardization, *ISO 26142:2010: Hydrogen detection apparatus – Stationary applications*, 2010. Standard.

[17] L. Piccioni Costa, M. Guerreiro, E. Puchta, Y. Tadano, T. Antonini Alves, M. Kaster, and H. Siqueira, *Multilayer Perceptron*, p. 105. IEEE Computational Intelligence Society Open Book, 01 2023.

DOI 10.5162/iCCC2026/P18

[18] T. Hübert and U. Banach, “Response time of hydrogen sensors,” in *Proceedings of the Fifth International Conference on Hydrogen Safety*, pp. 9–11, 2013.

DOI 10.5162/iCCC2026/P19

Passive MEMS-Sensoren für das Condition Monitoring

Philip Schmitt[1,2], Martin Hoffmann[1]
[1]Ruhr-Universität Bochum, Lehrstuhl für Mikrosystemtechnik
[2]mechIC GmbH, Bochum
Martin.hoffmann-mst@rub.de

Kurzfassung

Die Überwachung von Zuständen von Bauteilen in komplexen Anlagen ist heute vor allem begrenzt durch den Aufwand, den die Energieversorgung einer lokaler Sensorik verursacht, weil bislang weitgehend elektrische Energie notwendig ist, um Parameter zu messen und zu speichern. Viele Prozesse oder Ereignisse bringen Energie mit, für die Erzeugung elektrischer Energie durch Harvesting reicht es oft aber nicht. Hier können mikromechanische Sensoren, die vollständig passiv Ereignisse in unterschiedlichen Domänen (thermisch, mechanisch, elektrische oder magnetische Felder) bewerten und speichern und die Information in einer später digital abrufen Form abspeichern, eine Alternative sein. Einige Konzepte hierzu werden vorgestellt.

Abstract

The surveillance of complex systems is mainly limited by the energy supply. In almost all cases electrical energy is required for measuring and storage of data. While many processes inherently contain energy, energy harvesting is nonetheless still challenging. MEMS-bases fully passive sensor an record events in many domains (thermal, mechanical, electric or magnetic fields) and store the information in a digital format mechanically. Some concepts are reported here.

1 Die Idee passiver Sensoren

Die vorbeugende Instandhaltung profitiert von der frühzeitigen Information über entweder langsam ablaufende Alterungen oder über ungewöhnliche Belastungen im Alltagsbetrieb. Beide Informationen können mit Sensoren aus dem laufende Betrieb gewonnen werden. Allerdings fallen dabei große Datenmengen an, die keine Information enthalten, und die Extraktion relevanter Information erfordert permanent Energie.
Energy Harvesting aus der Umwelt für diese Hilfsenergie gelingt nur in manchen Fällen, während mit elektrisch vollständig passive Sensoren unter Nutzung der Mikromechanik eine Klasse neuer Sensorsysteme für die Maschinenüberwachung möglich wird: Sie misst, bewertet und speichert kontinuierlich und mechanisch, stellt das Ergebnis aber in digitaler elektrischer Form bereit, z.B. zur Auslesung mittels RFID Readern / Smartphones [1, 2].

2 Systemkonzept

In den letzten Jahren ist eine Toolbox entstanden, die durch Kombination einfacher MEMS Sensorknoten ermöglicht, die wahlweise Extremwerte erfassen oder die Anzahl von Grenzwert-Überschreitungen zählen. Das Kernstück dieser Sensorknoten ist ein mikromechanischer AD-Wandler, der die Verschiebung eines Balkens in einen binären Code überträgt und dessen Abtastung mittels einfacher Kontakte ermöglicht. Er ist verbunden mit einem Rast-Speicher, der primäre analoge Signale in diskrete Rastzustände übersetzt. Dies kann der jeweils größte / kleinste Wert einer Amplitude sein oder die inkrementelle Erhöhung eines Zählers, wenn ein Grenzwert überschritten wird. Die Energie, die für den Speichervorgang benötigt wird, stammt aus der zu messenden, meist nicht-elektrischen Größe; elektrische Hilfsenergie im Sensor wird für all diese Vorgänge nicht benötigt. Ein wichtiges mechanisches Element in dieser Kette ist ein Hebelsystem zur „Impedanzanpassung“: der Speicher braucht einen minimalen Weg und eine minimale Kraft, um die Rasten zu schalten. Meist reicht die Energie, aber das Verhältnis aus Weg und Kraft passt nicht. Hier erlaubt der Hebel eine effiziente Anpassung.

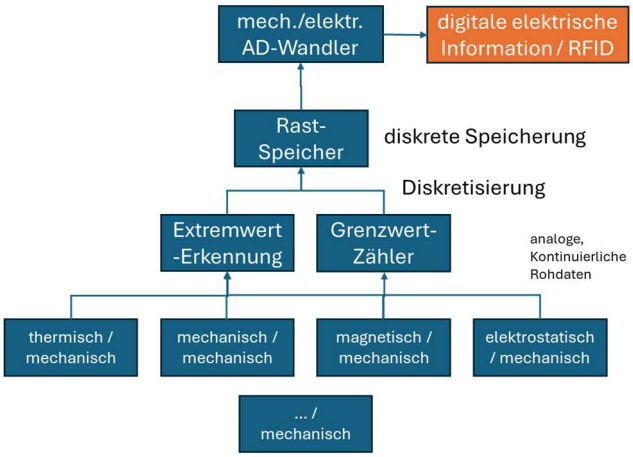


Bild 1 Systemkonzept eines passiven Sensorknotens mit den Hierarchie-Ebenen vom primären analogen Messwert zur diskreten, gespeicherten Information. Oben links: realisiertes RFID-Sensortag mit Stoßsensor

Bild 2 realisiertes RFID-Sensortag mit Stoßsensor

DOI 10.5162/iCCC2026/P19

3 Mikromechanische Komponenten

Zur Realisierung einer mechanischen Messkette werden mechanische Komponenten benötigt, die Funktionen wie Verstärkung, Datenselektion, Speicherung und Digitalisierung ermöglichen. In diesem Zusammenhang dient die Kraft als Analogie zum elektrischen Strom, die statische Verschiebung entspricht dem elektrischen Potential.

3.1 Energiequelle

Innerhalb der mechanischen Messkette kann die erforderliche Energie direkt aus der Messgröße entnommen und in mechanische Energie, also das Produkt aus Kraft und Weg, umgewandelt werden. Ein typischer MEMS-Energiewandler kann daher als lineare Kraftquelle betrachtet werden. Analog zu einer Stromquelle liefert die mechanische Quelle eine maximale Kraft, die von der angeschlossenen Last abhängt und durch die interne mechanische Impedanz begrenzt wird, welche im statischen Fall durch die interne Steifigkeit beschrieben wird. Ein mechanischer Kurzschluss der Quelle würde somit zu einer maximalen Auslenkung führen, die durch die interne Steifigkeit begrenzt ist.

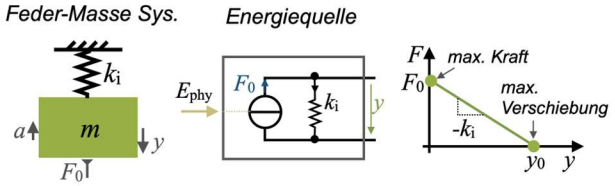
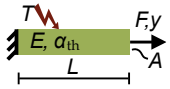
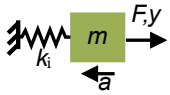
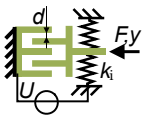


Bild 3 MEMS-Energiewandler als Quelle

Die folgende Tabelle 1 zeigt einige Beispiele für Energiewandler, die als äquivalente Quelle betrachtet werden können.

Tabelle 1: Energiewandler Übersicht

MEMS-Energiewandler	max. Kraft F_0	max. Verschiebung. y_0	inter. Steifig. k_i
Thermisch 	$A\,E\,\alpha_{th}\,\Delta T$	$L\cdot\alpha_{th}\,\Delta T$	$\frac{AE}{L}$
Inertial 	$m\,a$	$\frac{1}{k_i} m\,a$	k_i
Elektrostatisch 	$\varepsilon_0 \frac{n\,b}{d}\,U^2$	$\frac{1}{k_i}\varepsilon_0 \frac{n\,b}{d}\,U^2$	k_i

Auch indirekte MEMS-Energiewandler können mechanische Energie liefern. So ist etwa die Messung von Magnetfeldern oder elektrischer Ströme mit Hilfe von in Silicium integrierter Mikromagnete möglich. Hierzu lassen sich wie in [6] gezeigt, Mikromagnete in Pulverform direkt in SOI-MEMS integrieren. Abb. 4 zeigt den konzeptionellen Ansatz. Während die Messergebnisse in Abbildung 5 zeigen, dass mittels der integrierten Magnete hohe Kräfte und Verschiebungen für MEMS-Anwendungen generierbar sind.

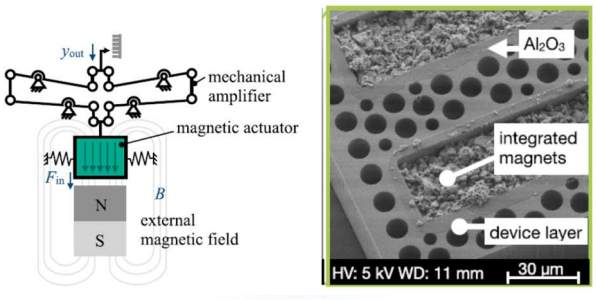


Bild 4 Magnetfeld-Sensor: Konzept u. Realisierung [5]

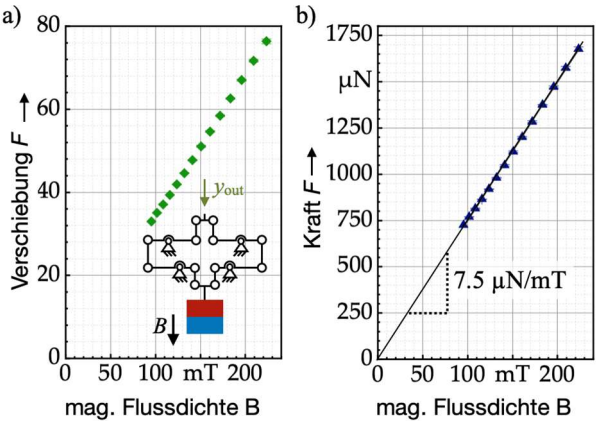


Bild 5 Exemplarische Messung der generierten Kraft und Verschiebung in Abhängigkeit der Flussdichte an einem MEMS-Magnetfeld Wandler

3.2 Transformatoren

Abhängig vom verwendeten Wandler und dessen interner Steifigkeit wird das primäre physikalische Signal entweder in eine große Kraft bei kleiner Auslenkung oder umgekehrt in eine große Auslenkung bei kleiner Kraft umgesetzt. Analog zu elektrischen Netzwerken, in denen Spannungen gegenüber Strömen häufig die bevorzugten primären Analogsignale darstellen, ist in mechanischen Netzwerken die Verarbeitung von Verschiebungen meist praktikabler als die Verarbeitung von Kräften. Daher werden Kräfte mithilfe eines hebelbasierten mechanischen Transformator in Verschiebungen umgewandelt; dieser wirkt hier als mechanischer Impedanzwandler.
Das Funktionsprinzip sowie der zugehörige Vierpol sind in Abb. 4 dargestellt. Das Verstärkungsverhältnis ist definiert als Verhältnis der Ausgangsverschiebung zur Eingangsverschiebung. Durch geeignete Auslegung der Hebelsegmentlängen (vgl. Abb. 4) ergibt sich näherungsweise

$$A_{\text{total}} = \frac{y_{\text{out}}}{y_{\text{in}}} = \frac{F_{\text{in}}}{F_{\text{out}}} \approx \frac{c}{b}\frac{f}{e} \qquad (1)$$

Es konnten bisher konstante Verstärkungsfaktoren von bis zu etwa 200 realisiert werden [3].

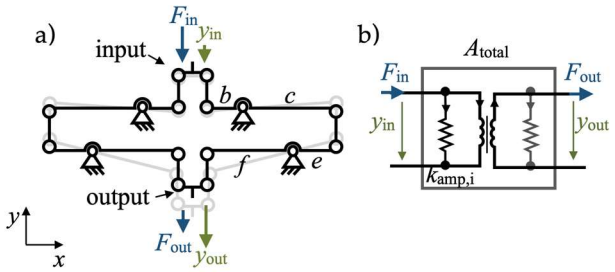


Bild 6 a) Mechanischer Transformator und b) Vierpol-Ersatzschaltbild

3.3 Speicher

Um ausgewählte Grenzwerte zu speichern, erfordert es einen mikromechanischen Speicher, der etwa den erreichten Maximalwert fixiert. Rastzahnspeicher eigenen sich hierfür besonders gut. Ein Verschiebungssignal, das vom Transformator bereitgestellt wird, soll über Rastzähne fixiert werden. Jedoch besteht die Herausforderung darin, dass die Fixierung des Läufers keinen Einfluss auf die folgenden Messungen haben soll. Von daher eignet sich die Entkopplung zwischen Transformator und Speicher über eine Kupplung. Dies entspricht dabei etwa der Funktion einer Diode, Verschiebungen werden nur in eine Richtung zugelassen. Die Kombination aus Kupplung und Speicher, wie in Abb. 7 gezeigt, ermöglicht es, dass jeweils nur der nächste größere Maximalwert gespeichert wird.

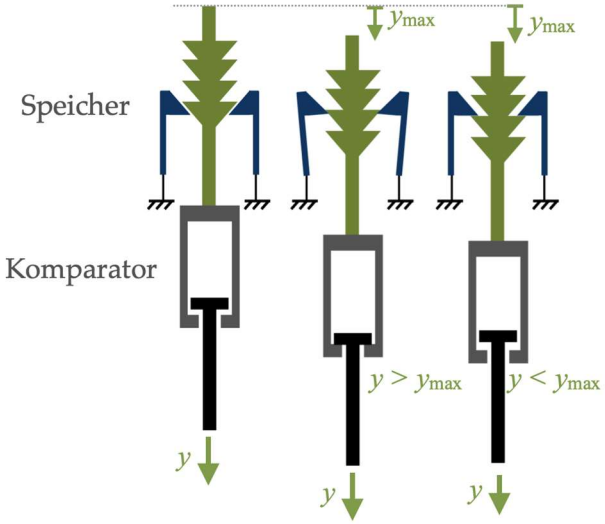


Bild 7 Maximalwertspeicher

3.4 Analog-Digital Wandler

Um den mechanischen Speicherzustand digital auszulesen, wird ein mechanischer A/D-Wandler (mechADC) verwendet. Der mechADC arbeitet als Positionsencoder, der die gespeicherte Verschiebung aus dem mechanischen Speicher in einen binären Dual- oder Gray-Code umwandelt. Der Encoder umfasst parallele Bit-Zweige mit Fingerelektroden, die den Binärcode geometrisch festlegen. Diese Bit-Zweige sind mit dem Speicherausgang verbunden und können in y-Richtung verschoben werden. Gegenelektroden, die fest am Rahmen angebracht sind, werden parallel zu den beweglichen Bit-Zweigen angeordnet, sodass sie beim Verschieben Kontakt mit den Fingerelektroden herstellen. Finger- und Gegenelektroden wirken zusammen als elektrische Schalter und ermöglichen eine binäre Positionskodierung.

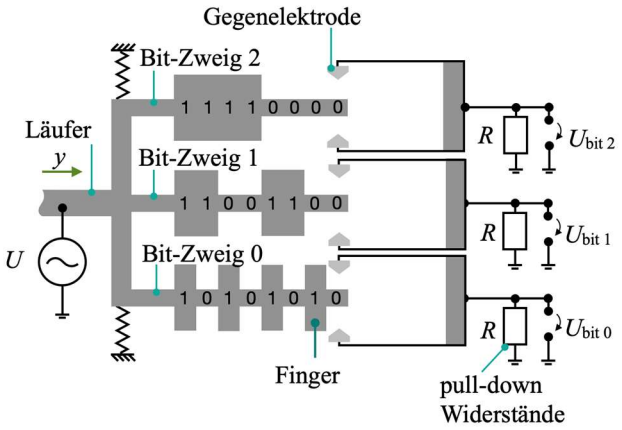


Bild 8 Prinzip des mechanischen A/D Wandlers

4 Beispiele für realisierte Systeme

4.1 Kraft- bzw. Stoßsensor

Die einfachste Grundkonfiguration ist eine federgelagerte Masse, die bei einer Einwirkung einer Beschleunigung oder eines Stoßes über eine Kupplung die maximale Auslenkung entsprechend der maximal erzeugten Kraft auf den Rastmechanismus überträgt: ist die aktuelle Kraft größer als ein zuvor erreichter Wert, wird der Rastmechanismus weiter ausgelenkt und der neue Wert wird festgehalten [4]. Von großer Bedeutung ist hier eine hochselektive Lagerung der Masse, so dass die Messung nur in der gewünschten Richtung erfolgt. Zu diesem Zweck wurden hochselektive Federmechanismen entworfen, deren Kennlinie in einem weiten Bereich einstellbar ist.

4.2 Rücksetzbare Sensoren

Da mechanische Rastmechanismen nur in eine Richtung Vorgänge erfassen, nach z. B. einem Auslesen im Rahmen einer Inspektion aber der Sensor auch rücksetzbar sein soll, können elektrostatische Aktoren integriert werden, die mittels RFID-Befehl den Rastmechanismus kurzzeitig lösen und damit die gespeicherten Werte löschen. Die notwendige Spannung kann problemlos aus der Reader-Feldenergie gewonnen werden. Die Zuverlässigkeit der Rastmechanismen wie der AD-Wandler wurde bislang bereits für viele tausend Schaltzyklen erprobt. [1, 2]

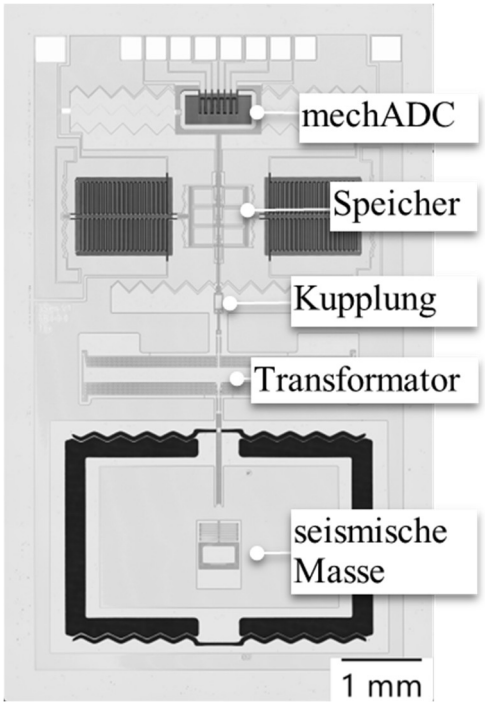


Bild 9 Stoß-/Kraftsensor mit Hebelmechanismus und mikromechanischem digitalem Encoder sowie elektrostatischem Aktor für das Rücksetzen

5 Zusammenfassung

In diesem Beitrag wird eine Toolbox für die vorbeugende Instandhaltung vorgestellt. Basierend auf den hier gezeigten Ideen hat sich das Start-Up mechIC (mechIC.de) ausgegründet, das ein erstes Sensorkonzept derzeit für die kommerzielle Nutzung umsetzt.
In der Zukunft ist geplant, die Toolbox um weitere Elemente zu erweitern und gemeinsam mit Nutzern konkrete Anwendungsszenarien zu erschließen. Die dafür notwendige Basistechnologie entspricht der für heute in Großserie verwendete MEMS-Sensoren und die Kopplung kann an bereits verfügbare RFID-Sensoren erfolgen.

6 Literatur

[1] M. Hoffmann, et al., Storing MEMS Interfaces Without Electrical Auxiliary Energy for Long-Time Monitoring, 2023 IEEE 36th International Conference on Micro Electro Mechanical Systems (MEMS), pp. 522-525 DOI: 10.1109/MEMS49605.2023.10052634

[2] P. Schmitt, M. Hoffmann, Sensors at the Edge: Towards Micromechanical in-Memory Computing, 2023 IEEE International Conference on Omni-layer Intelligent Systems (COINS), Berlin, Germany, 2023, pp. 1-5, DOI: 10.1109/COINS57856.2023.10189322

[3] P. Schmitt, M. Hoffmann, Engineering a compliant mechanical amplifier for MEMS-Sensor Applications, Journal of Microelectromechanical Systems, 2020. 29(2): S. 214-227 DOI: 10.1109/JMEMS.2020.2965260

[4] P. Schmitt, M. Hoffmann, A passive acceleration sensor with mechanical 6 bit memory and mechanical analog-to-digital converter, Micro and Nano Engineering, 2022, 100142, DOI: 10.1016/j.mne.2022.100142.

[5] P. Schmitt, M. Hoffmann, Direct Binary Encoding of Displacements on the Nano-Scale, 2020 IEEE 33rd Int. Conf. on Micro Electro Mechanical Systems, DOI: 10.1109/MEMS46641.2020.9056116.

[6] P.Schmitt, B. Gojdka, T. Lisec, M. Kroll, M.Hoffmann: SOI Integrated Micromagnets for mechanical Magnetic Field Detection IEEE Sensors Letters, 2023 DOI: 10.1109/LSENS.2023.3308125.

Diese Forschungsarbeiten wurden vom BMBF im Rahmen des ForMikro-Projektes UpFUSE gefördert.

DOI 10.5162/iCCC2026/P20

Umsetzbarkeit eines autarken solarbetriebenen HiFi-Lautsprechersystems

Antonio Arnold[1], Manuel Eberle[1], Thomas Stoppe[1], Frank Ellinger[1]
[1]TU Dresden, Chair for Circuit Design and Network Theory, Dresden, Germany
thomas.stoppe1@tu-dresden.de, Ellinger, Frank <frank.ellinger@tu-dresden.de>

Kurzfassung

Batteriebetriebene Geräte bieten eine mobile Stromversorgung für kleine und mittelgroße Geräte. Während mikroelektronische Geräte Strom im mW-Bereich benötigen, nutzen Leistungsverstärker für Lautsprecher bis zu mehrere hundert Watt. Auch wenn sie für den kabellosen Einsatz konzipiert werden können, muss die Batterie dennoch irgendwann aufgeladen werden. Wir haben die Möglichkeit untersucht, einen HiFi-Lautsprecher unter Verwendung von monokristallinen Solarmodulen mit einer Batterie als Energiespeicher zu betreiben. Es wurde die Wirksamkeit der Solarmodule unter verschiedenen Randbedingungen im Innenbereich untersucht und bewertet. Der Leistungsbedarf der Schaltung beträgt ca. 2.31 W, was bei 10 Stunden Ladezeit unter optimalen Bedingungen einen zuverlässigen Betrieb von 1 Stunde ermöglicht.

Abstract

Battery powered devices provide mobile power supply for small and medium devices. While microelectronic devices need power in the mW range, power amplifiers for loudspeakers require up to several hundreds of Watts. Even though they can be designed for wireless use, loading the battery is still required at some point.
We investigated the possibility of operating a hi-fi speaker using monocrystalline solar modules with a battery as energy storage. The effectiveness of the solar modules was investigated and evaluated under various boundary conditions indoors. The circuit's power consumption is approximately 2.31 W, which allows for 1 hour of reliable operation with 10 hours of charging time under optimal conditions.

1 Hintergrund

Während mikroelektronische Geräte Strom im mW-Bereich benötigen, nutzen Leistungsverstärker für Lautsprecher bis zu mehreren hundert Watt. Im Sinne einer nachhaltigen Energieversorgung ist zu klären, ob heute kommerziell verfügbare Energie Harvesting-Methoden für einen dauerhaften Betrieb und in Kombination mit einer Batterie sinnvoll dafür einsetzbar sind. Es existieren kommerzielle Lösungen für batteriebetriebene Lautsprechersysteme, die über Solar-Panele geladen werden können.
Der französische Hersteller PikiP bietet das Modell BOOTH an, das Frequenzen von 40 bis 20.000 Hz abdeckt und einen Schalldruckpegel von 102 dB bei 115 W erreicht. Mit einem Akkumulator von 2160 Wh kann es bis zu 45 Stunden ohne externe Energie betrieben werden; eine integrierte Photovoltaikanlage mit 320 W lädt den Energiespeicher in 6 Stunden auf.
Eine tragbare Alternative ist der Malibu von Urbanista, der Frequenzen von 20 bis 20.000 Hz abdeckt und 82 dB bei 10 W erreicht. Sein Akkumulator mit 13,32 Wh ermöglicht einen Betrieb von bis zu 20 Stunden. Die Photovoltaikeinheit arbeitet am effizientesten bei direkter Sonneneinstrahlung.
Die ATB 5 von WHD dient als Informationssystem. Sie hat einen Frequenzbereich von 400 bis 13.000 Hz und erreicht 97 dB bei 5 W Leistung, mit einer Laufzeit von bis zu 7 Stunden. Spezifikationen zur Photovoltaikanlage sind nicht frei verfügbar.

Um den maximalen Wirkungsgrad aus solarbetriebenen Systemen zu ziehen, sind diese idealerweise für die Nutzung im Freien konzipiert. Die Nutzung von Photovoltaik in Innenräumen ist stark vom genutzten Material (z. B. Silizium oder organische Halbleiter) und damit der Effizienz abhängig. Modelle wie der Malibu nutzen solche neuartigen Solarzellen, die auch unter künstlichem Licht funktionieren, jedoch ist die Ladeeffizienz in Innenräumen aufgrund der geringen Lichtintensitäten und des lichtquellenabhängigen optischen Spektrums limitiert.
Da räumliche Gegebenheiten wie Positionierung einen Einfluss auf die Klangqualität haben, soll untersucht werden, wie Ausrichtung und Abschattung die autarke Nutzung von autarken HiFi-Boxen beeinflussen. Zusätzlich soll die Systembetrachtung der elektronischen Komponenten Erkenntnisse für eine Optimierung eines solchen Systems liefern.

2 Materialien

Systementwurf

Die Wahl eines geeigneten Lautsprechers (elektromechanischer Wandler) mit gleichmäßigem Klangspektrum und passendem Verstärker für eine gute Klangqualität bei geringem Leistungsverbrauch bildet den Ausgangspunkt für die Gesamtbetrachtung des Systems.
Mit Blick auf die Entwicklung eines autarken Systems, wurden im Projekt weitere ingenieurstechnische Design-Aspekte betrachtet: Es wird Energy-Harvesting mit marktgängigen Solar-Modulen eingesetzt. Weiterhin sind Si-

DOI 10.5162/iCCC2026/P20

cherheitsaspekte betrachtet worden, da der generelle Einsatz von Akkus größerer Kapazität bei falscher Handhabung (z. B. Überladung, thermische Überlastung) ein Sicherheitsrisiko darstellt. Dies wurde durch Verwendung eines $LiFePO_4$-Akkus und eines Maximum Power Point Tracking (MPPT)-Reglers (siehe nachfolgender Absatz) mit entsprechenden Sicherheitsschaltkreisen realisiert.

MPPT-Regler
Als Solarladeregler wird der BlueSolar MPPT 75/10 des Herstellers Victron Energy eingesetzt. Dieser Laderegler verwendet einen besonders schnellen MPPT-Algorithmus, der im Vergleich zu konventionellen Ladereglern einen bis zu 10 % höheren Leistungsertrag ermöglichen kann. Weiterhin verfügt dieses Modell über integrierte Schutzfunktionen, die zur Erhöhung der Betriebssicherheit der Lautsprecherbox beitragen. Der Laderegler erfüllt die Normen EN/IEC 62109-1, UL 1741 und CSA C22.2.

Hi-Fi Lautsprecher
Zur effizienten Nutzung der verfügbaren Energie wird der Visaton B 200 Breitbandlautsprecher genutzt. Mit einem Schalldruckpegel von 96 dB (2,83 V/ 1 m) und einem Frequenzbereich von 40 Hz bis 18.000 Hz ermöglicht er eine energieeffiziente und akustisch hochwertige Klangwiedergabe.

Batterie
Die Energiespeicherung erfolgt über einen $LiFePO_4$-Akkumulator mit 128 Wh. Ein integriertes Batteriemanagementsystem sorgt für einen sicheren Betrieb, indem es den Akku vor normrelevanten Fehlzuständen schützt, u.a.:
- Kurzschluss
- Überladen
- Tiefentladung
- Überlastung (z. B. thermisch).

Zusätzlich ist die verwendete $LiFePO_4$-Zellchemie weniger anfällig gegen thermisches Durchgehen als Li-Ionen-Technologie.

Solar-Panele
Zur Deckung des Energiebedarfs wurden zwei monokristalline Solarzellen des Typs Sun Plus 10 S des Herstellers Phaesun eingesetzt. Die Module verfügen über eine Nennleistung von je 10 W.

Verschaltung
Das autarke Lautsprechersystem wurde zur Konzeptionierung in drei Teilsysteme untergliedert: die Photovoltaik, den Energiespeicher und die Verbraucher. Die Verschaltung dieser Teilsystemen mit den jeweiligen Komponenten ist in Bild 1 dargestellt.

Die Photovoltaikeinheit umfasst zwei monokristalline Solarmodule sowie den verwendeten MPPT-Ladereglers. Die Solarmodule werden in Reihenschaltung betrieben, da der Laderegler eine Eingangsspannung fordert, die mindestens 5 V über der Batteriespannung liegt.

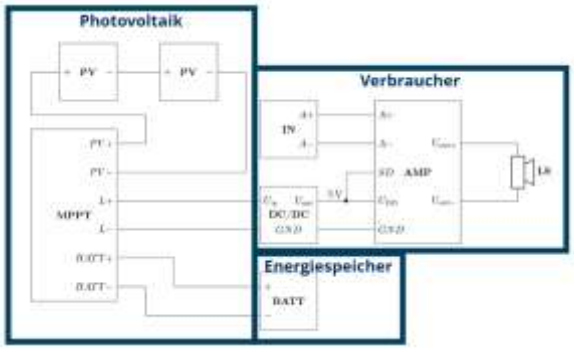


Bild 1 Verschaltung der autarken Lautsprecherbox

Am Laderegler sind Anschlüsse für den Energiespeicher sowie ein Lastausgang vorgesehen. Der ausgewählte Energiespeicher kann direkt an den Ausgang des Ladereglers angeschlossen werden.

Der Lautsprecher wird über einen Audioverstärker sowohl mit dem Audiosignal als auch mit der erforderlichen elektrischen Leistung versorgt. Zwischen Audioverstärker und Lastausgang des MPPT-Ladereglers ist ein DC/DC-Wandler erforderlich, um die Spannung von 12 V auf 5 V zu reduzieren.

3 Messungen

3.1 Messaufbau und Messort

Es wurden Messreihen durchgeführt, um die Leistungserzeugung der monokristallinen Solarmodule unter Innenraumbedingungen bewerten zu können.

Bild 2 Gemessene Leistung der Solarmodule im Innenraum in Abhängigkeit von der Ausrichtung

In Bild 2 ist die erzeugte Leistung in Abhängigkeit der Ausrichtung der Fensterfront des Lautsprecher-Raumes abgebildet. Alle Messungen erfolgten bei geschlossenen Fenstern auf dem Fensterbrett. Die Leistung ist bei süd- und westseitiger Ausrichtung vergleichbar hoch. Zudem zeigte sich, dass die Leistung mit zunehmender Bewölkung abnimmt. Wie erwartet, wird bei nördlicher Ausrichtung eine geringere Leistung erzielt. Auffällig ist jedoch, dass auf der Nordseite unter bewölkten Bedingungen eine höhere Leistung gemessen wird als bei sonnigen Bedingungen. Dies lässt sich durch Reflexionen des Sonnenlichts an den Wolken erklären, wodurch ein größerer Anteil diffusen

DOI 10.5162/iCCC2026/P20

und reflektierten Lichts vom Solarmodul absorbiert werden kann.

3.2 Einfluss Verschattung

Weiterhin wurde der Einfluss von Verschattungen auf die Leistungsabgabe der monokristallinen Solarmodule untersucht. Zunächst wurden gezielt einzelne Zellreihen des Solarmoduls abgedeckt.

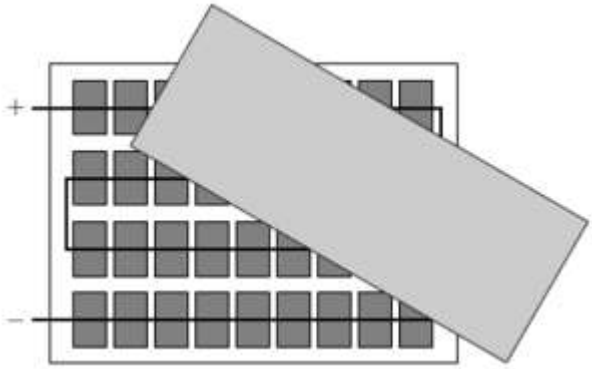


Bild 4 Unregelmäßige Verschattung des Solarmoduls

Bei der gezielten Verschattung einzelner Modulreihen von zwei parallel betriebenen Paneelen bleibt das Leistungsverhältnis nahezu konstant. Im Durchschnitt wird bei gezielter Verschattung (eine Reihe, zwei Reihen oder ein ganzes Modul) im Durchschnitt 37,60 % der Leistung ohne Verschattung erreicht. Ein eindeutiger Zusammenhang zwischen der verschatteten Fläche und dem Leistungsverhältnis lässt sich somit nicht erkennen.

Auch der Einfluss unregelmäßiger Verschattungen wurde untersucht. Diese sind schematisch in Bild 4 für die Fläche eines halben Panels abgebildet. Dabei werden auch diagonale Bereiche abgedeckt.
Die unregelmäßige Verschattung von einem halben Panel (entspricht der Fläche von zwei Reihen), in Reihenschaltung, führt zu einem deutlich geringeren Leistungsverhältnis von durchschnittlich 3,16 %. In der Reihenschaltung wirkt sich ein Abfallen des Stroms wesentlich stärker limitierend auf die Leistung des unverschatteten zweiten Moduls aus als der Spannungsabfall in der Parallelschaltung.

3.3 Einfluss Wetterbedingungen

Bild 5 Leistung in Abhängigkeit der Wetterbedingungen für zwei Solarmodule im Innenbereich

Bild 5 zeigt die erzeugte Leistung in Abhängigkeit der Wetterbedingungen. Sämtliche Messungen wurden bei südlicher oder westlicher Ausrichtung des Solarmoduls durchgeführt. Es zeigt sich eine klare Abhängigkeit: Mit zunehmender Bewölkung nimmt die Leistung der Solarmodule ab. Dieses Verhalten tritt sowohl bei geöffneten als auch geschlossenen Fenstern auf. Unter sonnigen Bedingungen und geöffneten Fenstern wird eine maximale Leistung zwischen 9,294 W und maximal 13,103 W bei zwei in Reihe geschalteten Solarmodulen erreicht.

4 Evaluierung Autarkes HiFi-System

Die entwickelte Lautsprecherbox (Bild 6) stellt im aktuellen Stand der Technik eine einzigartige Lösung dar, die keine vergleichbaren Produkte auf dem Markt hat.
Die Anforderung nach einem Schalldruckpegel (SPL) von mindestens 95 dB wurde nahezu erreicht, was die Effektivität der Lautsprecherbox unterstreicht.

Dennoch sind die vorgestellte Lautsprecherbox sowie die untersuchte kommerziell verfügbare Hardware nur eingeschränkt für den Einsatz im Innenraum geeignet. Zukünftige Fortschritte im Bereich der Materialforschung (organischen Photovoltaik - OPV, der Farbstoffsolarzellen - DSSC) mit größerer Effizienz würde zusätzliche Vorteile durch die Verwendung von Indoor-Spektren erlauben und damit eine kombinierte Verwendung ermöglichen, was zu einer verbesserten und zuverlässigeren Energieversorgung führt.

Der ermittelte Leistungsbedarf der Schaltung im Betrieb beträgt ca. 2,31 W, sodass bei einstündiger Nutzungsdauer des Systems bei einer Mindestladezeit von 10 h (durchschnittlich 700 mW Ladeleistung abzüglich Verluste durch Regler und Batterie) unter günstigen Wetterbedingungen ein zuverlässiger Betrieb erfolgen kann.

Bild 6 Lautsprecherboxsystem

DOI 10.5162/iCCC2026/P20

5 Ausblick

Anhand der Ergebnisse der experimentellen Untersuchungen wird erkennbar, dass zwei Modifikationen der ursprünglichen Schaltung erforderlich sind: Zum einen sollten bei einer Reihenschaltung der Solarmodule Bypass-Dioden integriert werden, um Leistungsverluste infolge von Verschattungen zu reduzieren. Zum anderen zeigt die gesamtenergetische Betrachtung, dass beim MPPT-Laderegler Optimierungsbedarf besteht.
Für den Einsatz in diesem System fällt eine hohe Verlustleistung von 300 mW an diesem an. Durch eine Schaltung zur Unterbrechung des Energieverbrauchs des Reglers bei Nichtnutzung der Ladefunktion kann der Standby-Energieverbrauch stark reduziert werden (Bild 7). Die Dimensionierung für kleine Leistungen birgt zudem Potential für eine Miniaturisierung.

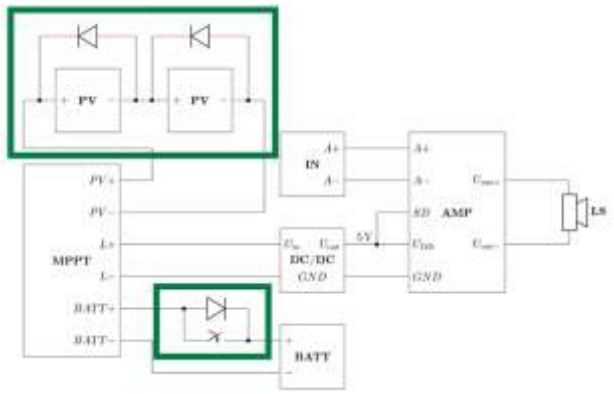


Bild 7 Anpassungen der Schaltung

Die Verwendung amorpher Solarzellen zur Deckung des Energiebedarfs unter künstlicher Beleuchtung (Typs AM-8801 von Panasonic) wurde theoretisch betrachtet. Diese Module verfügen über eine Nennleistung von 196 mW. Aufgrund der amorphen Zelltechnologie ist unter künstlicher Beleuchtung ein höherer Wirkungsgrad im Vergleich zu monokristallinen Solarzellen zu erwarten.
Eine erste Recherche zeigt, dass eine höhere Leistungsklasse aktuell nicht verfügbar ist. Eine Skalierbarkeit mit mehreren Einzelzellen ist aufgrund des hohen Preises derzeit vergleichsweise nicht rentabel.

Mit zunehmender Distanz nimmt der Anteil des direkt durch das Fenster einfallenden und vom Solarmodul absorbierten Lichts, und damit die Leistung, ab. In einer ersten Messung zeigte, dass sich auch der Einfluss eines geöffneten Fensters mit zunehmender Entfernung in ähnlicher Weise verringert. Dies sollte in weiteren Experimenten untersucht werden.

Weiterhin kann ein drahtloser Datenempfänger in die Lautsprecherbox integriert werden, um eine flexible und benutzerfreundliche Steuerung zu gewährleisten. Der Energiebedarf muss hierbei berücksichtig werden. Low-Power Empfänger sind für einen minimalen Standby-Strom prädestiniert.

Diese Konnektivität ist ebenso für den Aufbau einer zweiten Lautsprecherbox zur Erstellung eines kabellosen autarken Stereo-Paares relevant. Neben der energetischen Optimierung ist das System erst dann für die sinnvolle Anwendung im HiFi-Bereich (Stereo-Signale) relevant.

6 Danksagung

Beim DFG (Deutsche Forschungsgemeinschaft) Exzellenzcluster „Responsible Electronics in the Climate Change Era – REC2“ (EXC 3035, Projekt-ID 533607596) bedanken wir uns für den Wissensaustausch bzgl. Energie-Harvesting. Das Projekt „EKI-Saxony“ (100734135) wird mit Mitteln der Europäischen Union kofinanziert und durch Steuermittel auf Grundlage des vom Sächsischen Landtags beschlossenen Haushaltes mitfinanziert.

7 Literatur

[1] Adolf Goetzberger und Volker U. Hoffmann. Photovoltaic Solar Energy Generation (Springer Series in Optical Sciences). Springer, 2005, S. 232. isbn: 9783540236764.

[2] GuangyuWang. Technology, manufacturing and grid connection of photovoltaic solar cells. Hoboken, NJ, [Beijing], 2018.

[3] Fadliondi Fadliondi, Haris Isyanto und Budiyanto Budiyanto. „Bypass Diodes for Improving Solar Panel Performance“. In: International Journal of Electrical and Computer Engineering (IJECE) 8.5 (Okt. 2018), S. 2703. issn: 2088-8708. doi: 10.11591/ijece.v8i5.pp2703-2708.

[4] Visaton. Visaton B 200. 2023. url: https://www.visaton.de/sites/default/files/dd_product/B%20200_1350.pdf (besucht am 10. 06. 2025).

[5] Diodes Incorporated. Diodes Incorporated PAM8302A. 2021. url: https://www.diodes.com/datasheet/download/PAM8302A.pdf (besucht am 10. 06. 2025).

[6] Michael Safoutin u. a. „Effect of Current and SOC on Round-Trip Energy Efficiency of a Lithium-Iron Phosphate (LiFePO4) Battery Pack“. In: Bd. 2015. Apr. 2015. doi: 10.4271/2015-01-1186.

[7] Victron Energy. Victron Energy BlueSolar MPPT75/10. url: https://www.victronenergy.com/upload/documents/Datasheet-Blue-Solar-Charge-Controller-MPPT-75-10,-75-15,-100-15,-100-20_48V-DE.pdf (besucht am 10. 06. 2025).

[8] MEAN WELL. MEAN WELL N78 Series. 2024. url: https://www.meanwell.com/webapp/product/search.aspx?prod=N78 (besucht am 10. 06. 2025).

DOI 10.5162/iCCC2026/P21

Public Communication on the Energy Transition - a Natural Language Processing Approach for Analyzing News Data

Mareike Petrosjan, Dinesh Korrapati, Christin Hoffmann, Felix Müsgens, BTU Cottbus-Senftenberg, Cottbus, Germany
petrosja@b-tu.de, korradin@b-tu.de, hoffmanc@b-tu.de, felix.muesgens@b-tu.de

Abstract

Global mitigation targets as defined by the Paris Agreement can only be achieved through a fast and efficient energy transition. Pace and efficiency of the energy transition are strongly determined by its public acceptance. However, as the climate crisis itself, as well as one of its main driving solutions - the energy transition- is a complex, transformative, and politically diverse topic, communication about it is at risk of knowledge gaps, misinformation, and emotional polarization. Individual preferences are likely influenced and updated by any additional information received. Consequently, neutral and informative communication about the energy transition is becoming increasingly important, particularly in fostering public acceptance. We apply natural language processing (NLP) techniques to analyse public communication on the energy transition, focusing on sentiment analysis of transcripts from Germany's most-watched news program, *Tagesschau*. We address the challenge of efficiently detecting tone and emotion in this domain-specific context by offering a fine-tuned model that enables sentiment classification of texts related to the energy transition.

1 Introduction

Climate change constitutes one of the most pressing environmental, economic, and societal challenges of the present era [1]. Addressing mitigation targets, policymakers worldwide develop energy strategies [2], which encompass the large-scale adoption of renewable energy, improvements in energy efficiency, and the related adjustments in energy distribution, storage, and consumption patterns, collectively referred to as the energy transition [3]. Internationally, Germany is recognized as a pioneer in restructuring its energy system to incorporate an increasing share of renewable energy [4], [5]. To address the associated challenges, public acceptance plays a crucial role in the successful and rapid deployment of emerging energy technologies [6]. The transition demands substantial investment, labour market shifts, landscape interventions and broader societal transformation [7].

Given the complexity and transformative nature of the energy transition, substantial knowledge gaps exist, and individual preferences are continually shaped by new information received [8]. However, an efficient and effective energy transition depends on public acceptance that can be fostered by clear, neutral, and effective communication. [9]. On the contrary, limited understanding and emotional polarization may hinder well-informed decisions [10] that result in higher costs and more time needed for the realization of this transformation project.

Within the field of communication, the media plays an important role. Public opinion, political priorities, and media narratives influence each other [11]. While the media closely reacts to political debates [12], it can also elevate specific issues by selecting topics and framing them in particular ways, thereby shaping public attention and opinions [13]. In the context of the energy transition, media coverage can influence whether emerging technologies are facilitated or constrained [14].

Previous studies have mainly focused on newspaper portrayals of the German energy transition, examining aspects as scenario modelling communication [15] or international differences in reporting [16]. Applying topic modeling and sentiment analysis on coverage of the Renewable Energy Act (EEG), [2] could show a shift from positive to cost-focused critical framings. However, broader analyses beyond newspapers in Germany remain limited.

We focus our analysis on the *Tagesschau* as a high-reach news source that is accessed via traditional television broadcasting and digital devices. , recognizing its role as a key medium for information acquisition [17] and as a channel for science and environmental communication [18]. Our analysis is focusing on the energy transition media coverage of Germany's leading TV news program *Tagesschau* [19]. Some studies have analysed TV content related to aspects or specific dimensions of the energy transition. [20] analyse the climate change coverage in *Tagesschau* using neural topic modelling. For instance, [17] focuses on the TV news coverage of wind energy on three different American news channels.

In this study, we contribute to the literature by (1) addressing the research gap in the topic of public communication about the energy transition providing a sentiment analysis of the German TV format *Tagesschau* and (2) introducing a novel domain-specific model specialized on energy-related text data. Summarizing, three guiding questions structure our research:

- How is the energy transition communicated in the German news media debate over time?
- Which topics related to the energy transition dominate the news media discourse in Germany?

• Are there patterns of sentiment regarding specific topics of energy-related news?

For analysing large text corpora, Natural Language Processing (NLP) techniques have proven to be an effective and economical method compared to manual content analysis [21]. In order to better understand the emotional content of texts, sentiment analysis is a classical tool of use [22], which we will apply to our dataset.

2 Method

To understand the public communication about the energy transition, we analyse transcripts of the German newscast *Tagesschau*, which were prepared for natural language processing. We developed the domain-specific sentiment model E3M through a human–AI annotation workflow and a stacking ensemble approach. An overview of the NLP-Workflow guiding through the steps of pre-processing, filtering, sentiment analysis, and topic modelling is described in the following sections.

2.1 Tagesschau Data

The datasets used in this study consist of complete transcripts of the German news program *Tagesschau*. *Tagesschau* is a flagship newscast of Germany's public broadcasters, financed through mandatory household contributions, and is aired daily at 8 p.m. with a duration of 15 minutes and can be accessed online at any time. With an average audience of approximately 9.5 million viewers per day (as of 2024), it is the most-watched news program in Germany, with a market share of TV news of about 40.5 %. Among younger audiences, the program likewise maintains the highest viewership share [19]. A distinctive feature of *Tagesschau*, compared to international news formats, is its widespread perception as highly objective [23], [24].

2.2 Pre-processing

The dataset used in this study covers the period from October 30, 2007, to October 29, 2022, and continues with a second segment spanning July 12, 2023, to March 20, 2025. Program data were extracted from the publicly accessible *Tagesschau* archive using automated methods, with the first segment being provided through a dataset shared by [25]. For detailed analysis, the data were divided into one-minute segments. In a subsequent step, the audio tracks were extracted from the video files and automatically transcribed. For the second dataset, archived material containing program subtitles was accessed again.

The resulting raw dataset consists of unpunctuated transcribed speech. The absence of grammatical structures and the presence of transcription artefacts posed particular challenges for natural language processing. To address this, the "DeepMultilingualPunctuation" model [26] was applied during preprocessing to restore punctuation and grammatical structure, as it has demonstrated strong performance on German texts as well [27]. For subsequent topic delineation and sentiment analysis, the transcripts were segmented sentence by sentence and aligned with their respective timestamps, thereby increasing accuracy compared to sentiment values aggregated across full one-minute segments.

2.3 News Story Identification

To extract content relevant to the German energy transition, we applied a simple keyword-based approach comprising more than 200 terms. The keyword set covered domains such as energy sources, policy frameworks, and storage technologies, and was derived from publications of German energy agencies, specialized glossaries, journal articles, and newspapers. Following the keyword filtering, the challenge remained of ensuring that only thematically coherent and continuous reports were retained. To address this, we conducted manual checks for topic relevance and removed sentences that were not related to the topic. A translation into English using Python (Python Software Foundation, version 3.12.11) [28] was added to ensure that non-German speakers could contribute effectively to the process and to apply the models in the following procedure. Each news story, consisting of several sentences, was assigned a unique identification number per year to distinguish it from others during the subsequent steps.

2.4 Sentiment Analysis

The sentiment analysis in this study follows a stacking ensemble approach, wherein multiple base models are integrated through a supervised meta-learner that optimizes their combined predictive performance [29]. Prior research has demonstrated that hybrid deep-learning strategies can outperform single-model architectures by capturing complementary linguistic and contextual nuances across domains [30], [29]. Given the presence of both energy transition–related terminology and the distinct, predominantly objective tone of *Tagesschau* reporting, we expect that combining existing models would yield improved performance for our dataset.

2.4.1 Model Architecture

The developed sentiment framework follows a two-layer stacking structure. At the base layer, three transformer models are processing the translated dataset sentence by sentence: XLM-RoBERTa, a multilingual model, enables robust cross-linguistic sentiment recognition [31]. Soleimanian Financial-RoBERTa, trained on financial and economic news, captures nuanced polarity shifts within market- and energy-related narratives [32]. CardiffNLP Twitter-RoBERTa is tailored to detect informal, opinion-rich expressions characteristic of social media discourse [33]. Each model produces a SoftMax probability distribution, which expresses the model's confidence over the sentiment categories (positive, negative and neutral), where the sum of probabilities equals one. Instead of relying on "hard" class labels (e.g., only the most likely class), the ensemble uses these probability vectors, preserving uncertainty and nuanced confidence information from each model. The outputs from the three models (each with three

class probabilities) are concatenated into a nine-dimensional feature vector. Probability-level fusion of this kind has been shown to enhance accuracy and reduce classification noise in hybrid sentiment architectures [30].
At the meta layer, a Logistic Regression classifier functions as the meta-learner, learning optimal weights and interactions among the base-model outputs to produce the final sentiment prediction. Owing to its linear structure and interpretability, logistic regression provides an effective and widely adopted approach for integrating probabilistic features within stacking ensemble frameworks [29].

2.4.2 Training and Testing Data

In the subsequent step, we prepared the training-testing dataset for the sentiment analysis. [34] demonstrated the high potential of ChatGPT for annotation tasks, showing that the model can outperform human annotators in text classification. Similarly, [35] highlighted the use of Natural Language Explanations (NLEs) to support annotation tasks in cases where classifications are not immediately clear. While NLEs can reinforce human judgment, they may also bias annotators if the model's decision is incorrect, which requires careful handling when employing ChatGPT as an annotation support tool.
In our human–AI annotation workflow for the training-testing data, researchers first conducted independent sentiment annotations. ChatGPT then generated parallel annotations (positive/negative/neutral) together with one-sentence explanations. The same prompt was repeatedly used to ensure the consistent reliability of the statements. Researchers were able to review their initial assessments, particularly in cases of uncertainty and revise them with reference to the AI-generated outputs.
The training-testing set was developed to evaluate the accuracy of the newly created stacking ensemble model. Several identified news stories of 2022 were classified and used here as a 70/30 training-testing file. That means, 70% of the sentences were used for training and 30% for testing the model.
All models were implemented using the Hugging Face Transformers library [36] and executed on GPU hardware under fixed random seeds, ensuring reproducibility [37]The resulting Energy Economic ensemble model, E3M, is designed explicitly for energy-related data and achieved an overall accuracy of 77.78 %, exceeding the average accuracy of the individual base models. In the final step, we are applying this sentiment analysis sentence by sentence to the complete corpus of energy-related content.

2.4.3 Story-Level Sentiment Aggregation

To obtain a sentiment representation at the story level, the individual sentence-level predictions by the E3M model are aggregated into an overall score for each story. Each sentence was first assigned a numerical sentiment value (positive = +1, neutral = 0, negative = -1), enabling quantitative aggregation. For each story s comprising n sentences, the overall story-level sentiment score $\bar{S}_s$ was computed as the arithmetic mean of its sentence-level sentiment values, where v_i represents the sentiment value of sentence i (–1, 0, or +1), and n is the total number of sentences in story s:

$$\bar{S}_s = \frac{1}{n}\sum_{i=1}^{n} v_i$$

The resulting continuous score $\bar{S}_s$ was then discretized into categorical labels using fixed thresholds to ensure interpretability:

$$\text{Story Sentiment} = \begin{cases} \text{Positive}, & \text{if } \bar{S}_s > 0.2 \\ \text{Neutral}, & \text{if } -0.2 \le \bar{S}_s \le 0.2 \\ \text{Negative}, & \text{if } \bar{S}_s < -0.2 \end{cases}$$

This mean-based aggregation captures the overall polarity of each story while minimizing the impact of sentence-level variations.

2.5 Topic Modelling

To examine both the content and temporal patterns within the text corpus, we apply techniques of automated content analysis, specifically semi-automatic topic modelling, as the next step. Topic modelling is a computational approach designed to uncover main themes within large text corpora [38]. In a recent study, [39] used LLMs to perform topic modeling on social-media discussions related to data-center energy consumption. In this study, keyword-based extraction was integrated with language-model-assisted topic generation to enhance thematic coherence and contextual precision [40].
Each news story in the data set was annotated with three keywords describing the main theme. Here, KeyBERT, a method using BERT embeddings - numerical representations that capture the contextual meaning and usage of words, was applied [41]. Next, the large language model (LLM) Mistral-Instruct model [42], a version of Mistral designed to follow natural language instructions, was prompted to produce a single short label (up to four words) summarizing each story. The prompt included the extracted keywords and the full story text. To ensure consistent and reproducible results, the generation process employed a fixed random seed, a numerical value that guarantees identical outcomes across runs, and deterministic decoding with a temperature of 0. Deterministic decoding means that the model always selects the most probable word at each step, producing the same output for identical inputs and thereby eliminating randomness. The temperature hyperparameter controls the randomness of language model outputs: lower values, such as 0, make the model select the most probable tokens deterministically, resulting in stable and repeatable text generation, while higher values introduce greater variability [37].
In the next stage, we manually consolidated the model-generated topics to produce a more coherent thematic structure. We first established a set of broad thematic categories by conducting an exploratory, year-by-year analysis of the material. Subsequently, we identified consistent subtopics

for each narrative to harmonize terminology and avoid divergent labels for conceptually similar themes. News stories that address multiple themes are classified by the topic that exerts the most decisive influence. Consequently, thematic overlaps may occur, for example, a predefined main topic can be subsumed as a subtopic under another main topic, when its presence within the news story is less than that of the other one. Different main topics can include similar subtopics.

3 Preliminary Results

For presenting preliminary results, we are focusing on the time frame from July 2023 until March 2025, with particular attention to the modelled topics of energy policy and energy costs.

Table 1 shows an overview of these two main topics, including their subtopics, story and sentence count. Within the category of Energy Policy, the number of reports was the highest for energy pricing and political debates, including each eight news stories, followed by legislative measures such as the Gaspreisbremse and Gebäudeenergiegesetz, and industry prices. Several themes, including CO_2 pricing, consumer prices, energy saving, cooperation, Energiepreisbremse, energy law, and international affairs, were each mentioned only once, indicating a comparatively limited presence in the overall reporting.

The number of reports covering the main topic of energy costs is the highest for industry prices, with a story count of 16. Consumer prices and energy pricing are mentioned in two stories, while CO_2 pricing, Energiepreisbremse, Gaspreisbremse and Hospital Financial Crisis are mentioned each one time. Table 1 is showing the sentence count of each subtopic adding specific weight on the story count.

Main Topic	Subtopics	Story Count	Sentence Count
Energy Policy	Energy Pricing	8	132
	Political Debate Energy	8	111
	Gebäudeenergiegesetz	2	60
	Industry Prices	2	37
	Gaspreisbremse	2	29
	Energiepreisbremse	1	28
	Energy Saving	1	12
	Resources	1	11
	Cooperation	1	8
	CO_2 Pricing	1	7
	Consumer Prices	1	5
	Energy Law	1	3
	International	1	2
Energy Costs	Industry Prices	16	204
	Consumer Prices	2	23
	Energy Pricing	2	13
	Gaspreisbremse	1	26
	Energiepreisbremse	1	13
	CO_2 Pricing	1	10
	Hospital Financial Crisis	1	7

Table 1 Overview of Main Topics, Subtopics, Story and Sentence Count (Data from July 2023 - March 2025)

Regarding sentiment analysis, Figure 1 presents a combined visualization of monthly sentiment and story frequency for energy policy and energy costs from 2023 to 2025. The x-axis displays the timeline in monthly intervals. The left y-axis represents the mean sentiment scores, ranging from negative to positive values, and is depicted using line and point markers for both thematic categories. The right y-axis indicates the number of news stories published per month, illustrated as grey and hatched bars for energy policy and energy costs, respectively. In the topic of energy policy, months with the highest number of reports can be found in August, September, and November 2023, as well as in February 2025. The mean sentiment oscillates around neutral values with intermittent positive and negative peaks. In contrast, sentiment related to energy costs is consistently more negative, with several months showing pronounced downturns. Months with the highest number of reports here are August and September 2023.

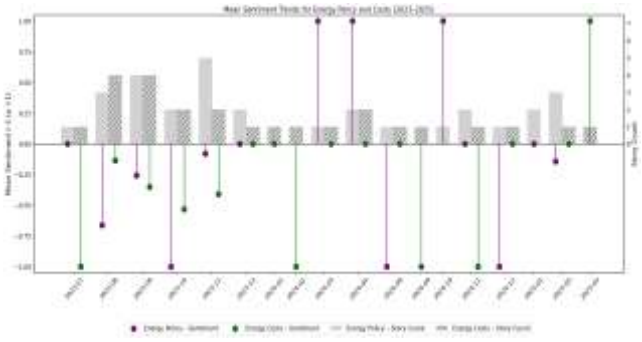


Figure 1 Sentiment Trends and Story Count on Energy Policy and Energy Costs

4 Discussion and Outlook

Our research approach offers particular value for scholars and data scientists applying quantitative methods in content analysis. In traditional manual approaches, assessing the emotional valence of topics through close reading is highly labour-intensive and time-consuming.

This study is subject to several limitations, which should be acknowledged. First, transcription errors may occur, but the large size of the dataset helps to mitigate the influence of individual inaccuracies. Second, sentiment analysis faces challenges in capturing subtle forms of expression such as implicit criticism, irony, or sarcasm; however, these cases are relatively infrequent in television news and thus unlikely to affect the broader patterns observed. Third, the reliance on television news data may introduce a slight bias toward neutrality, yet this also reflects the generally balanced reporting style of mainstream broadcasters. Finally, the analysis is restricted to a single national news channel, which ensures consistency and comparability; future research can extend the approach to regional channels for broader insights.

Acknowledgements

This research is supported by the German Federal Government, the Federal Ministry of Research, Technology and Space and the State of Brandenburg within the framework of the joint project EIZ: Energy Innovation Center (project numbers 85056897 and 03SF0693A) with funds from the Structural Development Act (Strukturstärkungsgesetz) for coal-mining regions. The paper benefited from helpful comments by participants of the Workshop on the Psychology of Climate Risk Communication Centre for Decision Research at the University of Leeds in September 2025. We thank our student assistants for their valuable support in conducting the manual content filtering.

5 Literaturverzeichnis

[1] R. C. Rial, „Biofuels versus climate change: Exploring potentials and challenges in the energy transition," Renewable and Sustainable Energy Reviews, Bd. 196, 2024.

[2] J. Dehler-Holland, K. Schumacher und W. Fichtner, „Topic Modeling Uncovers Shifts in Media Framing of the German Renewable Energy Act," Patterns, Bd. 2, Nr. 1, 2021.

[3] IPCC, „IPCC special report on renewable energy sources and climate change mitigation," Renewable Energy, Bd. 20, Nr. 11, 2011.

[4] R. Kunzig, „Energiewende—Vorbild Deutschland. In: National Geographic," 9 November 2017. [Online]. Available: https://nationalgeographic.de/umwelt/2017/11/energiewende-vorbild-deutschland/. [Zugriff am 22 September 2025].

[5] S. Strunz, „The German energy transition as a regime shift," Ecological Economics, Bd. 100, pp. 150-158, 2014.

[6] P. Emmerich, A.-G. Hülemeier, D. Jendryczko, M. J. Baumann, M. Weil und D. Baur, „Public acceptance of emerging energy technologies in context of the German energy transition," Energy Policy, Bd. 142, Nr. 111516, 2020.

[7] M. Pastukhova und K. Westphal, „Governing the global energy transformation," The geopolitics of the global energy transition, pp. 341-364, 2020.

[8] J. J. Häußermann, M. Maier und M. Schraudner, „Social acceptance of green hydrogen in Germany: building trust through responsible innovation," Energy, Sustainability and Society, Bd. 13, Nr. 22, 2023.

[9] D. Arlt, C. Schumann und J. Wolling, „What does the public know about technological solutions for achieving carbon neutrality? Citizens' knowledge of energy transition and the role of media.," Frontiers in communication, Bd. 8, Nr. 1005603, 2023.

[10] F. De Silva-Schmidt, Wissenserwerb zum Thema Klimapolitik aus Mediennutzung und interpersoneller Kommunikation, Bd. 31, Ilmenau: Universitätsverlag, 2021, pp. 617-633.

[11] M. E. McCombs, D. L. Shaw und D. H. Weaver, „New directions in agenda-setting theory and research," Mass Communication and Society, Bd. 17, pp. 781-802, 2014.

[12] W. L. Bennett, „Toward a Theory of Press-State," Journal of communication, Bd. 40, Nr. 2, pp. 103-127, 1990.

[13] T. Harcup und D. O'neill, „What is news? News values revisited (again)," Journalism studies, Bd. 18, Nr. 12, pp. 1470-1488, 2017.

[14] J. Lyytimäki, N. A. P. A. Nygrén und S. Rantala, „Energy transition looming behind the headlines? Newspaper coverage of biogas production in Finland," Energy, Sustainability and Society, Bd. 8, Nr. 1, p. 15, 2018.

[15] D. Scheer, „Communicating energy system modelling to the wider public: An analysis of German media coverage," Renewable and Sustainable Energy Reviews, Bd. 80, pp. 1389-1398, 2017.

[16] M. Antal und K. Karhunmaa, „The German energy transition in the British, Finnish and Hungarian news media," Nature Energy, Bd. 3, Nr. 11, pp. 994-1001, 2018.

[17] S. Gearhart, O. Adegbola und M. Guerra, „Harvesting the Wind: Analyzing Television News Coverage of Wind Energy," Environmental Communication, Bd. 13, Nr. 7, pp. 943-957, 2019.

[18] L. Kahlor und P. A. Stout, Communicating science, New York: Routledge, 2010.

[19] „NDR," Norddeutscher Rundfunk, 2 Januar 2025. [Online]. Available: https://www.ndr.de/der_ndr/presse/mitteilungen/Tagesschau-um-20-Uhr-erreicht-Spitzenwerte-und-bleibt-Deutschlands-meist-gesehene-Nachrichtensendung-,pressemeldungndr24888.html. [Zugriff am 19 September 2025].

[20] T. Schirmag, J. H. Wedemeyer und L. Wenz, „Neural topic modeling reveals German television's climate change coverage," Communications Earth & Environment, Bd. 6, Nr. 1, p. 441, 2025.

[21] S. Doropoulos, E. Karapalidou, P. Charitidis, S. Karakeva und S. Vologiannidis, „Beyond Manual Media Coding: Evaluating Large-Scale Text Analysis with NLP," Applied Sciences, Bd. 15, Nr. 14, 2025.

[22] B. Pang und L. Lee, „Opinion mining and sentiment analysis," Foundations and Trends® in information retrieval, Bd. 2, Nr. 1, pp. 1-135, 2008.

[23] „Häufige Fragen zur tagesschau: Gibt es eine Definition, wie eine ARD-aktuell-Sendung sein muss?," ARD, 2023. [Online]. Available: https://www.tagesschau.de/ueber-uns. [Zugriff am 22 September 2025].

[24] U. Paetzold, „Wie objektiv können Nachrichten sein,“ Gewerkschaftliche Monatshefte, Bd. 24, Nr. 8, pp. 488-495, 1973.

[25] R. Tschötschel, N. Schumann, R. Roloff und M. Brüggemann, „Der Klimawandel im öffentlich-rechtlichen Fernsehen. Inhaltsanalyse der "Tagesschau" und des Gesamtprogramms von Das Erste, ZDF und WDR 2007 bis 2022,“ Media Perspektiven, Bd. 51, Nr. 12, pp. 574-851, 2022.

[26] O. Guhr, A.-K. Schumann, F. Bahrmann und H.-J. Böhme, „Github,“ 26 September 2021. [Online]. Available: https://github.com/oliverguhr/deepmultilingualpunctuation. [Zugriff am 2025].

[27] O. Guhr, A.-K. Schumann, F. Bahrmann und H.-J. Böhme, „FullStop: Multilingual Deep Models for Punctuation Prediction,“ 2021.

[28] Python Software Foundation, „Python: A dynamic, open source programming language.,“ [Online]. Available: https://www.python.org/. [Zugriff am 2025].

[29] M. S. Sivri, „Combining sentiment analysis models using stacking ensemble learning techniques on BIST30 stocks,“ Hendese Teknik Bilimler ve Mühendislik Dergisi, Bd. 1, Nr. 2, pp. 91-97, 2024.

[30] K. L. Tan, C. Lee, K. Lim und K. Anbanathen, „Sentiment analysis with ensemble hybrid deep learning model,“ IEEE Access, Bd. 10, pp. 103694-103704, 2022.

[31] A. Conneau, A. Khandelwal, N. Goyal, V. Chaudhary, G. Wenzek, F. Guzman und V. Stoyanov, „Unsupervised cross-lingual representation learning at scale,“ Proceedings of the 58th annual meeting of the association for computational linguistics, pp. 8440-8451, 2020.

[32] M. Soleimanian, „Board Environmental, Social, and Governance (ESG) Expertise and the Usefulness of ESG Reports,“ 2024.

[33] F. Barbieri, J. Camacho-Collados, L. Anke und L. Neves, „TweetEval: Unified benchmark and comparative evaluation for tweet classification,“ Findings of the association for computational linguistics: EMNLP 2020, pp. 1644-1650, 2020.

[34] F. Gilardi, M. Alizadeh und K. Maël, „ChatGPT outperforms crowd workers for text-annotation tasks.,“ Proceedings of the National Academy of Sciences, Bd. 120, Nr. 30, 2023.

[35] F. Huang, H. Kwak und J. An, „Is chatgpt better than human annotators? potential and limitations of chatgpt in explaining implicit hate speech.,“ Companion proceedings of the ACM web conference, pp. 294-297, 2023.

[36] T. Wolf, L. Debut, V. Sanh, J. Chaumond und C. Delangue, „Huggingface's transformers: State-of-the-art natural language processing,“ arXiv preprint arXiv:1910.03771, 2019.

[37] A. Holtzman, J. Buys, L. Du, M. Forbes und Y. Choi, „The curious case of neural text degeneration,“ arXiv preprint arXiv:1904.09751, 2019.

[38] D. M. Blei, Y. Andrew und I. Michael, „Latent dirichlet allocation,“ Journal of machine Learning research, Bd. 3, pp. 993-1022, 2003.

[39] H. Wang, W. Hua, J. Peng und M. Hu, „Public sentiment analysis of data center energy consumption using social media data and large language models,“ Energy and Buildings , p. 115802, 2025.

[40] Y. Zhu, H. Yuan, S. Wang, J. Liu und C. Deng, „Large language models for information retrieval: A survey,“ ACM Transactions on Information Systems, 2023.

[41] M. Grootendorst, „KeyBERT: Minimal keyword extraction with BERT,“ 2020.

[42] A. Jiang, A. De Wynter, N. Lambert und C. Raffel, „The curious case of neural text degeneration,“ Proceedings of the International Conference on Learning Representations (ICLR), 2020.

[43] N. Masaling und D. Suhartono, „Utilizing RoBERTa and XLM-RoBERTa pre-trained model for structured sentiment analysis,“ INTERNATIONAL JOURNAL, Bd. 13, Nr. 3, p. 410, 2024.

[44] H. Kaplan, R.-P. Mundani, H. Rölke und A. Weichselbraun, „CrudeBERT: applying economic theory towards fine-tuning transformer-based sentiment analysis models to the crude oil market,“ arXiv preprint arXiv:2305.06140, 2023.

[45] P. Zhang, J. Rojcek und M. Leippold, „EconBERT: A Large Language Model for Economics,“ Available at SSRN 5263616, 2025.

[46] A. Conneau, K. Khandelwal, N. Goyal, V. Chaudhary, G. Wenzek, F. Guzmán, É. Grave, M. Ott, L. Zettlemoyer und V. Stoyan, „Facebook AI, Hugging Face,“ 26 September 2019. [Online]. Available: https://huggingface.co/FacebookAI/xlm-roberta-base. [Zugriff am 2025].

DOI 10.5162/iCCC2026/P22

Modellierung atmosphärischer Turbulenz zur erweiterten Simulation von schwankenden Windlasten auf Solar- und Windenergieanlagen

Modeling Atmospheric Turbulence for Advanced Simulations of Unsteady Wind Loads on Solar and Wind Energy Systems

Marten Klein

YIG Advanced Modeling and Simulation of Transfer Processes for Next Generation Energy Systems,
Scientific Computing (SC) Lab, Energie-Innovationszentrum (EIZ), Cottbus, Deutschland
Lehrstuhl Numerische Strömungs- und Gasdynamik, BTU Cottbus-Senftenberg, Cottbus, Deutschland
E-mail: marten.klein@b-tu.de

Kurzfassung

Der dauerhafte und effiziente Betrieb von Windenergie- und Photovoltaik- (PV-) Anlagen hängt maßgeblich davon ab, inwieweit die Anlagen wechselnden Windbedingungen standhalten. Sind Schäden erst einmal vorhanden, können sie sich im Laufe der Zeit durch wechselnde Lasten und Vibrationen, die durch Turbulenz und aerodynamische Instabilitäten angeregt werden, schnell verschlimmern. Extreme und wechselnde Windlasten verlässlich zu simulieren ist daher für die Bewertung der Lebensdauer von großer Bedeutung. Die vorliegende Arbeit demonstriert, wie Windlasten auf Solar- und Windenergiesysteme mithilfe verschiedener Softwarewerkzeuge simuliert werden können. Dabei liegt der Fokus auf der Modellierung der turbulenten Windverhältnisse als Voraussetzung für die Simulation variierender Lasten.

Abstract

Durable and efficient operation of wind turbines and photovoltaic (PV) systems depends on their ability to withstand highly variable wind conditions. Once initiated, damage can worsen quickly due to reversing loads and structural vibrations, excited by unsteady winds and aerodynamic instabilities. The ability to compute such unsteady wind loads is therefore important, for predesign, structural-health monitoring, and lifetime analysis. This paper demonstrates wind-load simulation capabilities across software tools, with applications to both solar and wind energy systems. Particular attention is paid to the modeling challenges and the representation of turbulent wind environments.

1 Introduction

Wind loads govern the lifetime and fatigue of structures and buildings. A detailed assessment, however, is rarely feasible due to limited availability of instrumented structures and computational resources. To overcome the associated design bottleneck, the long-standing paradigm in wind engineering has been to develop standardized rules for the design of structures and the wind loads associated to them based on specifically designed wind tunnel and field experiments. Relevant pioneering work can be attributed to Davenport, who introduced the concept of the wind loading chain [1] and advanced statistical descriptions on the basis of extreme value theory [2]. This approach has become established practice and has been formalized by wind loading standards and codes of practice (e. g. [3,4]). Regular updates, based on advances in research, reflect the specific needs of PV and wind energy systems.

Despite the progress made, challenges and limitations have remained with respect to unsteady loading and its consequences. PV systems (e. g. [5,6]) and wind turbines (e. g. [7,8]) are characterized by much smaller inertia than buildings and other structures due to their lightweight design in combination with a relatively large wind-exposed area and reliable loads obtained only experimentally. Computational approaches provide an economical alternative to costly experiments, offering further insight into the air flow, load and stress distributions, which are not easily accessible to measurements. However, modeling uncertainties can be significant, computational results in conflict with measurements or codes, resulting in disagreement on design provisions (e. g. [5]). For this reason, the main objective of this work is to showcase some modeling challenges associated with aerodynamics and atmospheric turbulence, affecting unsteady wind loading of PV and wind energy systems.

The rest of this paper is organized as follows. Section 2 introduces some fundamentals of computational wind engineering. Section 3 demonstrates computational approaches for the simulation of an onshore wind turbine and a ground-mounted PV system. Last, Section 4 closes with some concluding remarks.

2 Computational Wind Engineering in Wind and Solar Energy

Computational wind engineering addresses the development and application of simulations tools that provide a software implementation of Davenport's wind loading chain by a hierarchy of numerical models. Such tools can

DOI 10.5162/iCCC2026/P22

offer additional insight, but they are constrained by cost and fidelity. The highest fidelity is offered by computational fluid dynamics (CFD), which has predictive capabilities, but is also the most costly approach. CFD can provide detailed information based on approximate solutions to the governing conservation equations for mass, momentum, and energy (e. g. [6]). Engineering tools are less costly, but limited to lower fidelity due to utilization of reduced-order or surrogate models that rely on additional modeling assumptions (e. g. [8]). High- and low-fidelity simulation tools depend crucially on the representation of the local wind environment.

The wind environment of PV systems and wind turbines is governed by the atmospheric boundary layer (ABL) [9], which denotes the lower part of the atmosphere, up to height of about 1-2 km. The ABL exhibits characteristic wind fields that are influenced by terrain, vegetation, and surface properties (like texture and roughness), and energetically constrained by the local atmospheric conditions. Modeling of solar and wind energy systems is particularly challenged by atmospheric turbulence, which denotes the chaotic variability of atmospheric wind fields on time scales as short as seconds to minutes [11], calling for advanced modeling strategies.

3 Application Cases

3.1 Wind Turbine in Unsteady Conditions

Fast computation of unsteady wind loads benefits the development of intelligent control strategies for wind turbines in order to harmonize power generation, reduce fatigue loads, and steer wakes in wind farms (e. g. [8,12]). This requires reduced-order modeling of wind turbines, foremost rotor aerodynamics and generator torque, in response to the incoming wind of a turbulent ABL.

A representative example is given in **Figure 1**, in which a full-scale model of a DTU 10 MW wind turbine [13] is shown, computationally investigated with the aeroelastic wind turbine simulation tool QBlade [14]. The wind turbine has a hub height of 120 m and a rotor diameter of 178 m, composed of three blades. Aerodynamic forces are estimated for discrete segments of the structural beam elements based on simplifying assumptions, such as potential flow theory, and a wake model in order to represent the loss in wind speed due to wind power extraction and rotor-induced turbulence [8,14]. The structural response is visible by blade bending **(Figure 1a)**, here demonstrated for a mean wind speed of 10 m/s at hub height.

Turbulent atmospheric inflow is synthesized with TurbSim [15] by spectral filtering, which is one out of a number of available methods (e. g. [16,17,18]). TurbSim generates a 3-D turbulent box by spectral filtering of random numbers. The coarse spatial resolution (10 m distance between grid points in the inflow plane) can only represent large-scale flow features, like the red-colored wind gust in the inflow plane **(Figure 1b)**.

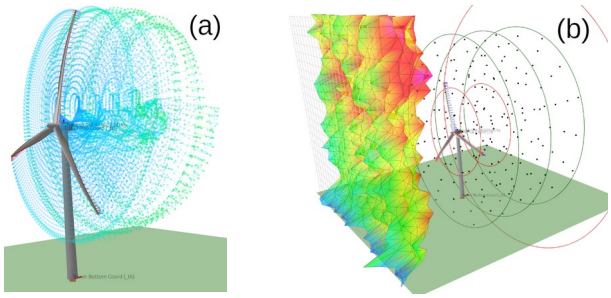


Figure 1 Dynamic simulation of a DTU 10 MW wind turbine with QBlade. (a) Snapshot showing the modeled free-vortex helical wake (blue) and blade bending. (b) Visualization of the instantaneous turbulent inflow as generated with TurbSim. Courtesy by Zubaer Haque, BTU.

A key limitation of spectral filtering approach is the construction of a “frozen-in” wind field. As suggested in [19], an alternative approach is the utilization of stochastic one-dimensional turbulence (ODT) [20], in a formulation applicable to the ABL [21,22]. ODT evolves the instantaneous wind velocity and additional scalar profiles if needed. Forcing and diffusive transport are directly resolved, whereas turbulent transport is modeled by a stochastic processes, formulated by mapping events, whose frequency depends on the locally available energy.

The modeled evolution of the instantaneous wind profile is shown in **Figure 2** for TurbSim, from the DTU 10 MW case in QBlade, and a canonical ABL case in ODT. The profiles shown represent snapshots taken at uniformly distributed output times, horizontally shifted by an arbitrary, constant offset. The profiles are normalized by the geostrophic free-stream velocity and the boundary-layer thickness, respectively. Present TurbSim results assume a parameterized reference wind profile with a hub-height velocity of 12 m/s at 120 m and a low-level jet of same speed at 100 m. Furthermore, IEC wind class Ia [23] properties are assumed, prescribing a turbulence intensity of 16%. In ODT, no feature parameterization is needed, since the solution emerges from the previously calibrated model [22] for the prescribed forcing, which is here a maintained nonzero free-stream velocity in geostrophic balance. The main difference between the two model setups is the nominally much lower wind speed in ODT to make full-scale vertical resolution numerically feasible.

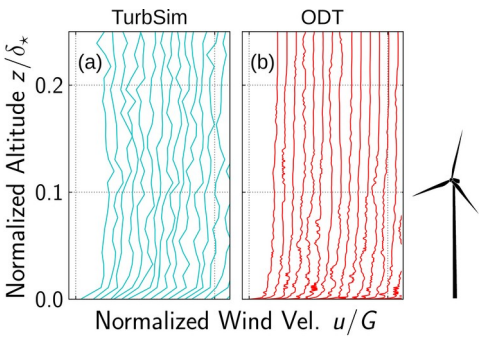


Figure 2 Instantaneous wind velocity profile evolution in (a) TurbSim (nominal Re = 1,222,000), and (b) standalone ODT (Re = 5,000). Profiles are normalized for comparison. A wind turbine is sketched as height reference.

DOI 10.5162/iCCC2026/P22

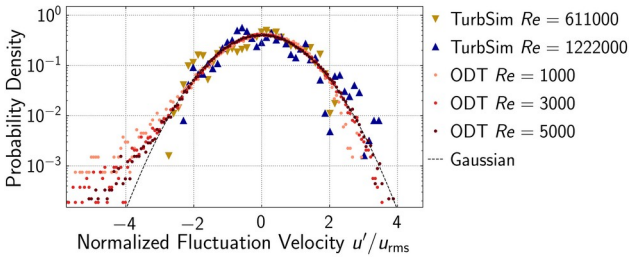


Figure 3 Probability density of the fluctuation velocity in the ABL. TurbSim cases correspond to a hub-height velocity of 6 m/s and 12 m/s, respectively. ODT simulations for various wind speeds at a lower magnitude regime exhibit saturation of non-Gaussian tails, indicating that wind gusts affect statistics beyond three standard deviations.

The velocity intermittency is shown in **Figure 3**, extending [19]. The fluctuation velocity is obtained from local time series and normalized by the standard deviation of the corresponding data record. The upside-down parabola in the semi-logarithmic axes represents a Gaussian normal distribution. Turbulent flows are intermittent and expected to deviate from the normal distribution, foremost in the tails (e. g. [18]). This property is reproduced by ODT for the turbulence-dominated region of the ABL. This holds even for fairly low wind speeds characterized by low Reynolds numbers (Re). Fluctuations in TurbSim are almost Gaussian, notably underestimating extreme wind gusts.

3.2 Unsteady Wind Loads on a PV System

Ground-mounted PV systems have a typical height of only a few meters and are hence located in the atmospheric surface layer [9]. The turbulence intensity increases nonlinearly towards the ground wheras the mean wind speed decreases (e. g. [1,2]). Turbulent wind gusts in the vicinity of the surface govern the unsteady wind loads, and the fatigue of panels and the mounting system. The low height of PV systems poses a challenge for wind tunnel measurements of such loads since the smaller scale models cannot be easily measured [24], rendering full-scale field measurements crucial. This challenge can be addressed computationally by high-fidelity CFD models, like the single-purpose CFD software PVade [25] and the multi-purpose CFD-library OpenFOAM [26]. In order to cut computational cost, a preliminary model comparison is performed for a minimalistic case with reduced computational burden by relaxing CFD best practice guidelines.
The minimalistic domain and configuration of the PV system investigated is given in **Figure 4**, showing a dual-axis solar tracker represented by a single square panel with 8 m edge length and a thickness of about 0.1 m. The mean velocity profile of the incoming wind, parameterized by a power law [2], is sketched on the left, at the inflow boundary of the flow domain. This profile is used to initialize the flow. Only the bottom boundary and PV panel are treated as no-slip boundaries. Top, front and back are free-slip, the left side inflow, and the right side outflow boundary conditions. Further details are given in [25].
An exemplary CFD solution obtained with PVade for low-wind conditions is shown in **Figure 5**. The central vertical slice shows the spatial distribution of the instantaneous velocity magnitude by blue-red pseudocolors after the initial transient as passed. The airflow separates at the panel edges and develops a turbulent wake downstream. The visible front half of the panel's sun-facing side visualizes the distribution of the instantaneous fluid stress that acts on the PV panel by another set of pseudocolors.
Last, simulated time series of the lift coefficient are shown in **Figure 6** for PVade and OpenFOAM. The two CFD solvers differ by their discretization schemes and subgrid-scale turbulence models, utilizing the Smagorinsky (Smag.) and the wall-adapting local eddy-viscosity (WALE) model. The initial transient is similar, but the saturation is different. OpenFOAM is generally more dissipative than PVade, damping down wind fluctuations and unsteady lift forces emerging due to vortex shedding.

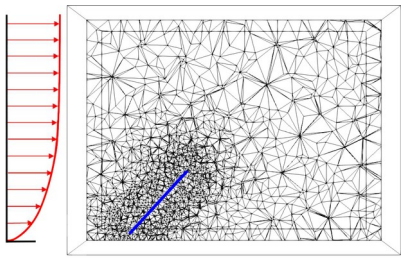


Figure 4 Unstructured mesh (black) of the flow domain around a ground-mounted square PV panel (blue) with an edge length of 8 m without mounting system. The incoming wind profile (red) is sketched and obeys a power law.

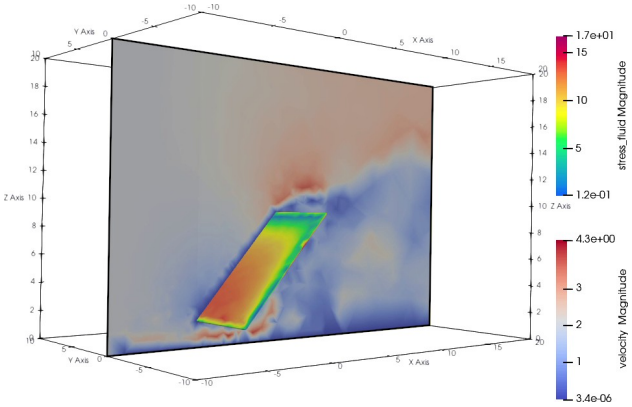


Figure 5 Visualization of the instantaneous wind field (blue-red) around a square PV system together with the wind stress (rainbow colors) obtained by a CFD simulation with PVade. Courtesy by Joseph Gerardi, BTU.

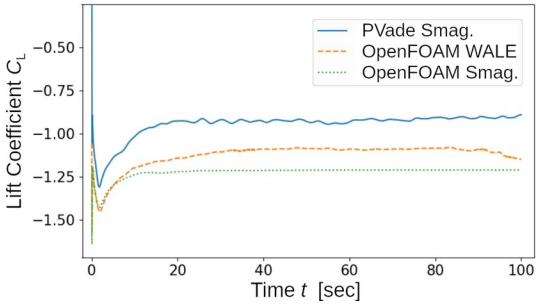


Figure 6 Simulated lift coefficient for a square PV panel varying the CFD solver and turbulence model. Courtesy by Hemanth N. Kotra, BTU.

4 Concluding Remarks

Numerical simulation of unsteady wind loads on PV and wind energy systems requires dynamical models capable of representing aerodynamic and turbulent flow features across scales and model fidelity. This is particularly the case for turbulent atmospheric boundary layers. Map-based stochastic modeling can serve as gap-bridging approach, addressing both cost and physical compatibility.

5 Acknowledgment

This research is supported by the German Federal Government, the Federal Ministry of Research, Technology and Space and the State of Brandenburg within the framework of the joint project EIZ: Energy Innovation Center (project no. 85056897, 03SF0693A) with funds from the Structural Development Act for coal-mining regions.

6 Literature

[1] Davenport, A. G.: A statistical approach to the treatment of wind loading on tall masts and suspension bridges. Ph. D. thesis. University of Bristol, 1961.

[2] Davenport, A. G.: Note on the distribution of the largest values of a random function with application to gust loading. In: Proc. Inst. Civil Eng. 28 (1964), p. 187.

[3] Eurocode 1: Actions on structures, Parts 1-4: General actions – Wind actions. EN 1991-1-4. Brussels: European Committee for Standardization, 2005.

[4] ASCE/SEI 7-05, -10, -16, -22: Minimum Design Loads for Buildings and Other Structures. Reston: American Society of Civil Engineers (ASCE), 2005, 2010, 2015, 2022.

[5] Stathopoulos, T.; Zisis, I.; Xypnitou, E.: Local and overall wind pressure and force coefficients for solar panels. J. Wind Eng. Ind. Aerodyn. 125 (2014), pp. 195-206.

[6] Wittwer, A. R.; Podestá, J. M.; Castro, H. G.; et al.: Wind loading and its effects on photovoltaic modules: An experimental–computational study to assess the stress on structure. Sol. Energ. 240 (2022), pp. 315-328.

[7] Veers, P.; Dykes, K.; Lantz, E.; et al.: Grand challenges in the science of wind energy. Science 366 (2019), 6464.

[8] Perez-Becker, S.: Advanced aerodynamic modeling and control strategies for load reduction in aeroelastic wind turbine simulations. Ph. D. thesis, Technische Universität Berlin, 2021.

[9] Garratt, J. R.: Review: the atmospheric boundary layer. Earth-Sci. Rev. 37 (1994), pp. 89-134.

[10] O'Malley, M.; Holttinen, H.; Cutululis, N.; et al.: Grand challenges of wind energy science – meeting the needs and services of the power system. Wind Energ. Sci. 9 (2024), pp. 2087-2112.

[11] Van der Hoven, I.: Power spectrum of horizontal wind speed in the frequency range from 0.0007 to 900 cycles per hour. J. Atmos. Sci. 14 (1957), p. 160-164.

[12] Sierra-Garcia, J. E.; Santos, M.: Deep learning and fuzzy logic to implement a hybrid wind turbine pitch control. Neural Comput. Applic. 34 (2022), pp. 10503-10517.

[13] Bak, C.; Zahle, F.; Bitsche, R.; et al.: Description of the DTU 10 MW reference wind turbine. DTU Wind Energy Report I-0092, Technical University of Denmark, 2013.

[14] Marten, D.: QBlade: A modern tool for the aeroelastic simulation of wind turbines. Ph. D. thesis, Technische Universität Berlin, 2019.

[15] Kelley, N. D.; Jonkman, B. J.: Overview of the TurbSim stochastic inflow turbulence simulator. Technical Report NREL/TP-500-41137, National Renewable Energy Laboratory, 2007.

[16] Mann, J.: Wind field simulation. Prob. Engng. Mech. 13 (1998), pp. 269-282.

[17] Quon, E. W.; Ghate, A. S.; Lele, S. K.: Enrichment methods for inflow turbulence generation in the atmospheric boundary layer. J. Phys.: Conf. Ser. 1037 (2018), 072054.

[18] Yassin, K.; Helms, A.; Moreno, D.; et al.: Applying a random time mapping to Mann-modeled turbulence for the generation of intermittent wind fields. Wind Energ. Sci. 8 (2023), pp. 1133-1152

[19] Klein, M.; Haque, Z.; Schmidt, H.: Map-based stochastic turbulence modeling for utilization in wind engineering. IFAC PapersOnLine 59-26 (2025), pp. 271-276.

[20] Kerstein, A. R.: One-dimensional turbulence: Model formulation and application to homogeneous turbulence, shear flows, and buoyant stratified flows. J. Fluid Mech. 392 (1999), pp. 277-334.

[21] Kerstein, A. R.; Wunsch, S.: Simulation of a stably stratified atmospheric boundary layer using one-dimensional turbulence. Bound.-Lay. Meteorol. 118 (2006), pp. 325-356.

[22] Klein, M.; Schmidt, H.: Exploring stratification effects in stable Ekman boundary layers using a stochastic one-dimensional turbulence model. Adv. Sci. Res. 19 (2022), pp. 117-136.

[23] IEC 61400-1: Wind Turbine Generator Systems, Part 1 – Safety Requirements. Geneva: International Electrotechnical Commission (IEC), 1994, 2019.

[24] Jubayer, C. M.; Hangan, H.: Numerical simulation of wind effects on a stand-alone ground mounted photovoltaic (PV) system. J. Wind Eng. Ind. Aerodyn. 134 (2014), pp. 56-64.

[25] Young, E.; He, X; King, R.; Corbus, D.: A fluid-structure interaction solver for investigating torsional galloping in solar-tracking photovoltaic panel arrays. J. Renew. Sustain. Energ. 12 (2020), 063503.

[26] Weller, H. G.; Tabor, G.; Jasak, H.; Fureby, C.: A tensorial approach to computational continuum mechanics using object-oriented techniques. Comput. Phys. 12 (1998), pp. 620-663.

DOI 10.5162/iCCC2026/P23

Auswirkungen von KI Training und Inferenz auf die Betriebskosten und das Abwärme Potential

Germain-Ray Schmidt, TU Berlin, FG Technologie und Management für Integrierte Energieinfrastrukturen und Geothermie, Straße des 17. Juni 135, 10623 Berlin, Deutschland, germain-ray.schmidt@campus.tu-berlin.de
Prof. Dr. Dr. Tanja Kneiske, TU Berlin, FG Technologie und Management für Integrierte Energieinfrastrukturen und Geothermie, Straße des 17. Juni 135, 10623 Berlin, Deutschland
Fraunhofer IEG, Gulbener Str. 23, 03046 Cottbus

1 Einleitung

Rechenzentren entwickeln sich zu einem der zentralen Treiber des weltweiten Stromverbrauchs. Der jüngste Sonderbericht der Internationalen Energieagentur (IEA) zu „Energy and AI“ schätzt den Strombedarf von Rechenzentren im Jahr 2024 auf rund 415 TWh und erwartet bis 2030 einen Anstieg auf etwa 945 TWh, wobei der Großteil dieses zusätzlichen Verbrauchs auf KI-Workloads und beschleunigte Server (GPUs, TPUs) zurückgeführt wird [1]. Für Deutschland zeigt die Studie „Rechenzentren 2022“, dass der Stromverbrauch deutscher Rechenzentren zwischen 2010 und 2022 von etwa 10 TWh auf 17,9 TWh gestiegen ist und dass gleichzeitig ein technisch nutzbares Abwärme potenzial von mehreren TWh pro Jahr besteht, das bislang nur punktuell genutzt wird [2]. Mit dem Energieeffizienzgesetz (EnEfG) reagiert der Gesetzgeber auf diese Entwicklung und verpflichtet Rechenzentren ab einer IT-Anschlussleistung von 300 kW zu höheren Effizienzstandards sowie zur Prüfung und schrittweisen Nutzung unvermeidbarer Abwärme. Der Bitkom-Leitfaden zum EnEfG fasst die Vorgaben zusammen und macht deutlich, dass neue Rechenzentren künftig Mindestanteile wiederverwendeter Energie nachweisen müssen und Abwärmenutzung damit zu einem verbindlichen Planungsparameter wird [3]. Parallel dazu verändert sich die technische Struktur der Rechenzentren grundlegend: Klassische, CPU-dominierte Infrastrukturen mit typischen Rack-Leistungsdichten von etwa 7–10 kW werden zunehmend durch KI-Cluster mit hohen Anteilen GPU-basierter Server ergänzt, die Leistungsdichten von 30 bis über 100 kW pro Rack erreichen und über lange Zeiträume hoch ausgelastet sind [4]. Experimentelle Messungen von Deep-Learning-Workloads zeigen, dass dabei der überwiegende Teil der IT-Energie in den GPUs umgesetzt wird und nahezu vollständig als Wärme im Rechenzentrum anfällt [5]. In Kombination mit flüssigkeitsbasierten Kühlsystemen eröffnen GPU-dichte KI-Rechenzentren damit neue Optionen für höherwertige Abwärmenutzung, stellen aber gleichzeitig neue Anforderungen an Betrieb, Wirtschaftlichkeit und Einbindung in Energie- und Wärmenetze [4] [6]. Vor diesem Hintergrund untersucht die vorliegende Arbeit, wie sich der Umstieg von CPU-dominierten auf GPU-basierte KI-Hardware auf die Betriebskostenstrukturen sowie das nutzbare Abwärmepotenzial von Rechenzentren unterschiedlicher Größenklassen (klein, mittel, groß, Hyperscale) auswirkt**.** Dazu werden ein CPU-Basisszenario mit überwiegend konventionellen IT-Servern und mehrere GPU-Szenarien mit unterschiedlichen Anteilen KI-beschleunigter Racks definiert. Auf Basis gemessener Leistungsdaten für Training und Inferenz sowie vereinfachter Annahmen zu Kühlung, Abwärmerückgewinnung und Energiepreisen werden die resultierenden Energieflüsse, Abwärmepotenziale und jährlichen Betriebskosten, sowohl mit als auch ohne Einspeisung der Rechenzentrumsabwärme vergleichend analysiert.

2 Hintergrund und Stand der Forschung

2.1 Entwicklung des Energiebedarfs und Rolle von KI

Der weltweite Energiebedarf von Rechenzentren wird in der aktuellen Literatur auf rund 300-380 TWh im Jahr 2023 geschätzt [6]. Parallel dazu zeigen Analysen zu KI Rechenzentren, dass der zusätzliche Strombedarf durch AI-Workloads bis 2030 auf 200-400 TWh anwachsen könnte, was dann 35-50 % des gesamten Rechenzentrumsstroms entspräche [6]. Damit wird deutlich, dass KI und damit GPU-beschleunigte Systeme künftig einen wesentlichen Anteil am Gesamtenergiebedarf der Rechenzentren ausmachen werden. Für Deutschland zeigt die Borderstep Studie „Rechenzentren 2022“, dass der Stromverbrauch der Rechenzentren zwischen 2010 und 2022 von etwa 10 TWh auf 17,9 TWh gestiegen ist. Davon entfielen 2022 rund 12 TWh auf IT- Hardware und knapp 6 TWh auf Infrastruktur wie Kühlung, USV und Stromverteilung [3]. Zugleich werden in dieser und nachfolgenden Analysen Abwärmepotenziale von mehreren TWh pro Jahr für Deutschland identifiziert, die bei geeigneter technischer und wirtschaftlicher Ausgestaltung in Wärmenetze eingespeist werden könnten [3]. Vor diesem Hintergrund und den gesetzlichen Anforderungen an Effizienz und Abwärmenutzung ergibt sich die Notwendigkeit, die energetischen Auswirkungen neuer KI- /GPU-Rechenzentren systematisch zu bewerten. 2.2 GPU-/KI-Rechenzentren und Leistungsprofile Der Übergang von klassischen CPU-Rechenzentren zu GPU-basierten KI-Rechenzentren verändert sowohl die Leistungsdichten als auch die Lastprofile. In konventionellen Rechenzentren werden Standardracks typischerweise mit 7-10 kW betrieben, während KI-Racks in spezialisierten KI-Rechenzentren 30-100+ kW pro Rack erreichen; in dedizierten KI-Facilities liegt der Durchschnitt über 60 kW pro Rack [1]. Moderne Hyperscale-KI-Rechenzentren wiesen Gesamtleistungen von >100 MW auf und werden teilweise in Richtung Gigawatt-Campi geplant [1]. Diese hohen Leistungsdichten stellen besondere Anforderungen

an Stromversorgung, Kühlung und Netzintegration. Auf Workload-Ebene unterscheiden sich KI-Lasten deutlich von klassischen IT-Workloads. Chen et al. beschreiben KI-Rechenzentren als durch hoch variable und „bursty“ Lastprofile geprägt, insbesondere in der Inferenzphase, während das Training großer Modelle über lange Zeit räume mit sehr hohen und relativ stabilen Leistungsaufnahmen läuft [1]. In einer Synthese aktueller Studien wird der Anteil der KI-Energie grob abgeschätzt auf etwa 60 % Inferenz, 30 % Training und 10 % Datenvorbereitung/Fine-Tuning [6]. Auf der Ebene einzelner KI-Server zeigen experimentelle Messungen von Aquino-Brítez et al., dass bei typischen Deep-Learning-Modellen (CNNs) der Großteil der IT- Energie von der GPU getragen wird: Während des Trainings entfallen rund 61-73 % der Energie auf die GPU, 14– 22 % auf die CPU und 13-17 % auf den Arbeitsspeicher. In der Inferenzphase liegt der GPU-Anteil immer noch bei 49-59%, CPU und RAM teilen sich den Rest [5]. Diese Ergebnisse liefern eine belastbare Grundlage, um leistungsbasierte Profile für GPU-Server zu definieren und für verschiedene Rechenzentrumsgrößen zu skalieren.

2.2 Abwärmepotenziale und Technologien zur Wärmerückgewinnung

Da nahezu die gesamte in Rechenzentren aufgenommene elektrische Energie letztlich in Wärme umgewandelt wird, stellen Rechenzentren konzentrierte Abwärmequellen dar. Hao et al. schätzen, dass typischerweise etwa 40 % des Stroms direkt in der IT-Hardware, weitere 40 % in den Kühlsystemen und die verbleibenden 20 % in sonstiger Infrastruktur umgesetzt werden und dass diese Energie nahezu vollständig als Niedertemperatur-Abwärme anfällt [2]. Die Nutzbarkeit dieser Abwärme hängt stark von der Kühltechnologie und dem resultierenden Temperaturniveau ab. Luftgekühlte Rechenzentren liefern üblicherweise Ablufttemperaturen im Bereich von etwa 25-40 °C, während flüssiggekühlte Systeme 40-80 °C erreichen können [2]. In einer umfassenden Bewertung von Abwärmerückgewinnungstechnologien schlagen Hao et al. einen exergiebasierten Effizienzindikator vor, der unterschiedliche Nutzungsoptionen) auf eine vergleichbare Basis „äquivalenter elektrischer Energie“ abbildet [2]. Die relativen Exergieeffizienzen liegen typischerweise im Bereich von etwa 7,5–11,5 % für direkte Heizungsanwendungen, 4,2-12,8 % für Kältebereitstellung und 3,0-13,2 % für Stromerzeugung. Diese Ergebnisse verdeutlichen, dass GPU-dichte KI Rechenzentren mit Flüssigkühlung nicht nur höhere spezifische Leistungen aufweisen, sondern aufgrund höherer Rücklauftemperaturen auch ein größeres nutzbares Abwärmepotenzial bieten können als klassische, luftgekühlte CPU-Rechenzentren. Genau an dieser Schnittstelle setzt die in dieser Arbeit entwickelte Szenario- und Methodikbetrachtung an.

3 Methodik

Die Arbeit verfolgt einen szenariobasierten**,** bottom-up Ansatz: Ausgehend von typisierten Rechenzentrumsgrößen werden CPU- und GPU-Lasten, Abwärmepotenziale und Betriebskosten für verschiedene Hardware-Mixe modelliert und miteinander verglichen. Die Methodik gliedert sich in fünf Schritte: Definition der Systemgrenzen und Rechenzentrumsgrößen, Szenariodefinition, Lastmodellierung von CPU- und GPU-Servern, Modellierung von PUE und Abwärmenutzung sowie Betriebskostenberechnung inkl. Sensitivitätsanalyse.

3.1 Systemgrenzen und Rechenzentrumsgrößen

Die Systemgrenze umfasst jeweils ein Rechenzentrum inklusive IT-Hardware (Server, Storage, Netzwerk) und technischer Infrastruktur (Stromversorgung, Kühlung, Nebenverbraucher). Vorketten (z. B. Herstellung der Hard ware) sowie nachgelagerte Netze (Fernwärmeverteilung) werden nicht bilanziert. Es werden vier idealtypische Rechenzentrumsgrößen betrachtet, die sich an der in Bitkom beschriebenen Strukturmodellen und Leistungsbereichen orientieren, aber um noch eine Hyperscale-Größe ergänzt wurde.[3]: kleines RZ mittleres RZ, großes RZ, Hyperscale-RZ.

Für jede Größenklasse werden eine typische Rackanzahl und eine mittlere IT-Leistung pro Rack festgelegt (kW/Rack). Diese Werte werden aus Bandbreiten für klassische und KI-Racks abgeleitet und so gewählt, dass sie im Rahmen der in [1], [3], [6] beschriebenen Größenordnungen liegen.

3.2 Szenariodefinition: CPU-Basisszenario und GPU-Szenarien

Für jede Rechenzentrumsgröße wird zunächst ein **CPU**-Basisszenario definiert: ausschließlicher Einsatz konventioneller IT-Server (CPU-dominierte Racks),ntypische Rack-Leistungsdichten klassischer Rechenzentren (8 kw/ Rack) überwiegend luftbasierte Kühlung, keine Abwärmenutzung

Darauf aufbauend werden mehrere GPU-Szenarien entwickelt, in denen der Anteil der GPU-basierten KI-Hardware an der gesamten IT-Leistung stufenweise steigt Die Szenarien sind so ausgelegt, dass: ein definierter Anteil klassischer IT-Server in allen Szenarien erhalten bleibt (nicht-KI-Workloads), die Gesamt-IT-Leistung der jeweiligen Größenklasse vergleichbar bleibt,

3.3 Lastmodellierung von CPU- und GPU-Servern

Die elektrische Leistungsaufnahme der IT-Komponenten wird getrennt für CPU- und GPU-Server modelliert.

CPU-Server

Für CPU-Server wird mangels detaillierter Lastprofile ein konstanter mittlerer Leistungswert pro Server angenommen, der eine realistische Jahresauslastung klassischer Rechenzentren abbildet. Dieser Wert wird aus Literaturangaben zu typischen Server- und Auslastungsniveaus abgeleitet [3], [6] und je

DOI 10.5162/iCCC2026/P23

Rechenzentrumsgröße skaliert. Zeitliche Schwankungen (z. B. Tagesgang) werden in dieser ersten Modellierung nicht explizit berücksichtigt; die Einflüsse werden in der Sensitivitätsanalyse diskutiert.

GPU-Server

Die Leistungsaufnahme der GPU-Server während Training und Inferenz wird aus experimentellen Messdaten von Aquino-Brítez et al. abgeleitet [5].

Aus den im Sensors-Paper angegebenen Energieverbräuchen und Laufzeiten für verschiedene Deep-Learning-Modelle auf NVIDIA-GPUs werden für jede Architektur mittlere Leistungen für Training Inferenz bestimmt. Für das Jahreslastprofil eines GPU-Servers wird ein vereinfachtes Duty-Cycle-Modell verwendet, das den Anteil von Training, Inferenz und Idle-Betrieb abbildet. Basierend auf Literaturangaben zum relativen Energieanteil von Training und Inferenz in KI-Rechenzentren (ca. 30 % Training, 60 % Inferenz, 10 % weitere KI-Aufgaben [1], [6]) werden für jedes Szenario

$$\alpha_{\text{train}}, \quad \alpha_{\text{infer}}, \quad \alpha_{\text{idle}}$$

die Zeitanteile
festgelegt. Die mittlere elektrische Leistung eines GPU - Servers ergibt sich zu

$$\bar{P}_{\text{GPU}} = \alpha_{\text{train}} P_{\text{train}} + \alpha_{\text{infer}} P_{\text{infer}} + \alpha_{\text{idle}} P_{\text{idle}}.$$

Unter der Annahme linearer Skalierung wird diese Serverleistung mit der Anzahl GPU-Server bzw. GPUs pro Rack multipliziert und so auf die jeweilige Rechenzentrumsgröße hochgerechnet. Die Gesamt-IT-Leistung ergibt sich aus der Summe der CPU- und GPU-Teilleistungen.

3.4 PUE, Abwärmepotenzial und Wärmerückgewinnung

Die energetische Infrastruktur wird über den Power Usage Effectiveness (PUE) modelliert. Für das CPU-Basisszenario wird je Rechenzentrum ein PUE Wert von 1,4 festgelegt. Für GPU-Szenarien mit höherer Leistungsdichte und Flüssigkühlung werden, im Einklang mit der Literatur zu modernen KI-Rechenzentren, reduzierte PUE-Werte im Bereich 1,1-1,3 angenommen [1], [2], [6]. Die Gesamtleistung des Rechenzentrums ergibt sich damit aus: $P_{\text{total}} = P_{\text{IT}} \cdot \text{PUE}$

Da die aufgenommene elektrische Energie im Rechenzent rum nahezu vollständig in Wärme umgewandelt wird, wird diese Leistung als Abwärmestrom interpretiert. Für die tatsächlich nutzbare Wärmemenge wird ein technologieabhängiger Wirkungsgrad der Wärmerückgewinnung eingeführt, der sowohl die Kühltechnik (Luft vs. Flüssig kühlung) als auch Verluste in Wärmeübertragern und Speichersystemen abbildet [2].

$$E_{\text{Wärme, nutzbar}} = P_{\text{IT}} \cdot \eta_{\text{WRG}} \cdot h_{\text{Betrieb}}$$

Für die Nutzung der Abwärme wird in dieser Arbeit eine Wärmepumpenlösung zur Einspeisung in ein Fernwärmenetz unterstellt, da diese in der Literatur als zentrale Option für die Kopplung von Rechenzentren und Wärmenetzen identifiziert wird [2]. Die dafür notwendige zusätzliche elektrische Energie der Wärmepumpe wird über angenommene Jahresarbeitszahlen modelliert.

3.5 Betriebskostenmodell und Sensitivitätsanalyse

Zur Bewertung der wirtschaftlichen Effekte wird ein vereinfachtes Betriebskostenmodell auf Basis jährlicher Energieflüsse verwendet. Investitionskosten (CapEx) der IT- und Kühltechnik werden in dieser ersten Betrachtung nicht detailliert bilanziert; der Fokus liegt auf den laufen den Energiekosten (OpEx) und möglichen Erlösen aus Wärmeeinspeisung.

Für jedes Szenario werden zunächst die jährlichen Stromkosten ohne Abwärmenutzung berechnet. Im Fall mit Abwärmenutzung werden zusätzlich die Stromkosten der Wärmepumpe berücksichtigt und Erlöse aus der Einspeisung der nutzbaren Wärme in ein Fernwärmenetz angesetzt. Aus der Differenz werden die relativen Kostenvorteile bzw. -nachteile der Abwärmenutzung für CPU- und GPU-Szenarien abgeleitet. Da zentrale Parameter (Strom- und Wärmepreise, PUE-Werte, Jahresarbeitszahlen) mit Unsicherheiten behaftet sind, ist eine Sensitivitätsanalyse vorgesehen. In dieser werden die genannten Größen in plausiblen Bandbreiten variiert und die Auswirkungen auf Abwärmepotenziale und Betriebskosten untersucht. Dies ermöglicht eine robuste Einordnung, ob und unter welchen Rahmenbedingungen GPU-basierte KI-Rechenzentren durch Abwärmenutzung wirtschaftliche Vorteile gegenüber CPU-dominierten Basisszenarien bieten können.

4 Erwartete Ergebnisse und Beitrag

Auf Basis der beschriebenen Szenarien wird für jede Rechenzentrumsgrößenklasse und jeden GPU-Anteil zu nächst die jährliche IT-Leistung und Gesamtenergie ermittelt. Es ist zu erwarten, dass mit zunehmendem GPU-An teil die absolute IT-Leistung und damit das theoretische Abwärmepotenzial deutlich steigen, insbesondere in den großen und Hyperscale-Szenarien. Gleichzeitig führen niedrigere PUE-Werte in flüssiggekühlten GPU-Szenarien dazu, dass der relative Anteil der IT-Leistung an der Gesamtleistung zunimmt, während der Energiebedarf der Infrastruktur anteilig sinkt. Für die Abwärmenutzung wird erwartet, dass GPU-dominierte Szenarien aufgrund höherer Leistungsdichten und höherer Rücklauftemperaturen der Kühlmedien größere nutzbare Wärmemengen für Fernwärmesysteme bereitstellen können als CPU-dominierte Basisszenarien. Insbesondere für große und Hyperscale-Rechenzentren dürften sich bei geeigneter Auslegung der Wärmerückgewinnung und ausreichend hoher Wärmenachfrage signifikante Einspeisemengen und damit

relevante Erlöspotenziale ergeben. Im Betriebskostenvergleich werden voraussichtlich zwei gegenläufige Effekte sichtbar: Einerseits führen höhere IT-Leistungen in GPU-Szenarien zu steigenden Stromkosten, andererseits können verbesserte PUE-Werte und Erlöse aus der Wärmeeinspeisung die spezifischen Betriebskosten pro IT-Leistung senken. Die Arbeit soll aufzeigen, in welchen Größenklassen und bei welchen GPU-Anteilen Abwärmenutzung dazu beitragen kann, die zusätzlichen Energiekosten von KI-Rechenzentren teilweise zu kompensieren. Der Beitrag der Untersuchung liegt damit in einem konsistenten Modellrahmen, der

(i) den Übergang von CPU- zu GPU-basierten Rechenzentren entlang verschiedener Größenklassen abbildet,

(ii) empirisch basierte Leistungsprofile von KI-Workloads integriert und

(iii) Energiebilanz, Abwärmepotenziale und Betriebskosten unter regulatorischen Rahmenbedingungen gemeinsam betrachtet. Dies liefert eine Grundlage für Planungs- und Standortentscheidungen von Betreibern sowie für energiewirtschaftliche Bewertungen von KI-Rechenzentren als Wärmequelle.

[1] IEA: Energy and AI, 2024

[2] Borderstep: Rechenzentren 2022, 2022

[3] Bitkom: Rechenzentren in Deutschland, 2024

[4] Chen, X.; Wang, X.; Colacelli, A.; Lee, M.; Xie, L.: Electricity Demand and Grid Impacts of AI Data Centers: Challenges and Prospects., 2025

[5] Aquino-Brítez, S.; García-Sánchez, P.; Ortiz, A.; Aquino-Brítez, D.: Towards an Energy Consumption Index for Deep Learning Models: A Comparative Analysis of Architectures, GPUs, and Measurement Tools., 2025 [6] IEA: Data Centre Energy Use: Critical Review of Models and Results, 2025

[7] Hao, Y.; Zhou, H.; Tian, T.; Zhang, W.; Zhou, X.; Shen,Datacenters waste heat recovery technologies: Review and evaluation.

DOI 10.5162/iCCC2026/P24

Effects of Surface Functionalization on Convective Heat Transfer with Relevance to Heat Exchanger Engineering

Sai Ravi Gupta Polasanapalli[1,2], Marten Klein[1,2], and Heiko Schmidt[1,2]

[1]Chair of Numerical Fluid and Gas Dynamics, BTU Cottbus-Senftenberg, 03046 Cottbus, Germany
[2]Scientific Computing Lab, Energie-Innovationszentrum (EIZ), BTU, Cottbus-Senftenberg, 03046 Cottbus, Germany
E-Mail-Address: SaiRaviGupta.Polasanapalli@b-tu.de

Abstract

This study investigates how wall surface properties affect fluid motion and heat transfer in an idealized natural convection setup. Natural convection in a cubic box with a heated bottom wall and a cooled top wall is used as a simplified setup to analyze heating and cooling efficiency for regular (no-slip), free-slip, and coated (finite-slip) walls. Using a three-dimensional lattice Boltzmann method (LBM) solver in direct numerical simulation (DNS) mode, simulations are performed for three Prandtl numbers Pr = 0.786, 4.38, and 10, and two Rayleigh numbers Ra = 2×10^6, and 10^7. The aim is to understand how surface conditions influence flow patterns, thermal mixing, and heat transfer efficiency. Results show that free-slip conditions significantly enhance heat transfer, yielding higher Nusselt numbers due to thinner thermal boundary layers and stronger convective currents. The results demonstrate that surface boundary conditions play a role in modulating flow dynamics and heat transfer in RB convection.

1 Introduction

Energy efficiency and effective thermal management are key challenges in modern energy systems, ranging from conventional power plants to renewable energy technologies. Improving heat transfer performance directly enhances system efficiency, reliability, and sustainability. One promising strategy to achieve this involves modifying the properties of solid surfaces in contact with fluids. In particular, hydrophobic and super-hydrophobic coatings have attracted attention because they can change how fluids move close to a wall, potentially reducing flow resistance [1] and improving heat transfer performance [2, 3]. Such surface effects are particularly important because even thin surface layers or coatings can significantly influence thermal performance. In practical heat exchangers, for example, the formation of thin rough material layers on internal surfaces can lead to rapid efficiency losses of about 10–15%, highlighting the strong sensitivity of heat transfer to surface conditions.

Understanding how wall conditions affect turbulent heat transfer is therefore crucial for the design and operation of energy-efficient thermal systems. In this context, Rayleigh–Bénard (RB) convection serves as a canonical model for buoyancy-driven heat transfer and is relevant to applications such as heat exchangers, solar thermal collectors, nuclear reactor cooling, and geothermal energy systems [4].

In this study, the natural convection in a box is studied as an idealized setup that enables quantitative analysis of heating/cooling efficiency under the influence of functionalized (coated) walls. Specifically, we examine no-slip, free-slip, and finite-slip boundary conditions with slip lengths in the range slip-lengths l_s = 0.00005m to 0.0005m. Three-dimensional numerical simulations are performed using a lattice Boltzmann method (LBM) solver, which is well-suited for efficient upscaling to high Rayleigh number regimes with strong thermal forcing. The LBM solver operates in direct numerical simulation (DNS) mode, ensuring full-scale resolution of the flow dynamics. The primary objective is to examine the effects of no-slip, free-slip, and finite-slip wall conditions on flow and heat transfer characteristics for fluids and chemical solvents with different Prandtl numbers. The results provide insights into optimizing thermal performance through surface control.

2 Simulation Setup

Figure 1(a) depicts the schematic representation of the RB cell under investigation. The computational domain is defined as H × H × H in the x, y, and z directions, where H represents the height of the cubic cavity, serving as the characteristic length scale. The bottom wall is at a higher temperature (T_h), while the top wall is at a lower temperature (T_c). Temperature boundary conditions on the vertical walls are considered as adiabatic walls. The flow and heat transfer in the RB system are governed by the Rayleigh and Prandtl numbers. The Rayleigh number represents the strength of buoyancy-driven convection relative to viscous and diffusive effects, while the Prandtl number describes the relative rates of momentum and thermal diffusion in the fluid. Together, these parameters determine the convection regime and heat transfer characteristics. In our study, we explore three distinct velocity boundary conditions, namely, no-slip, free-slip, and finite-slip, applied to all walls. The finite-slip condition is a commonly used mathematical framework for modelling hydrophobic and super-hydrophobic wall behaviour. It is given by Navier-slip is shown in **Figure 1(b)**. The Navier slip length (l_s), defined as the distance from the wall to the point where the extrapolated wall-parallel velocity component vanishes, is a standard choice

DOI 10.5162/iCCC2026/P24

[5]. A slip length of zero corresponds to a no-slip wall, while an infinite slip length represents a free-slip wall. Typical slip lengths range from 100 nm for hydrophobic surfaces to 100 μm for super-hydrophobic surfaces.

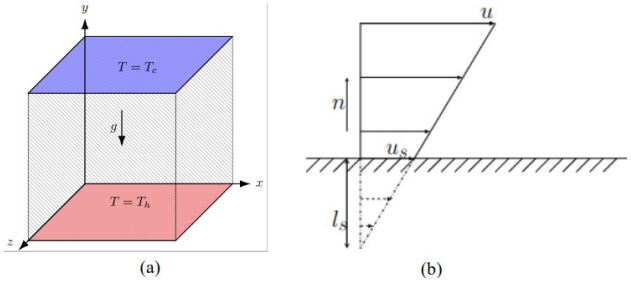


Figure 1 Schematic diagrams. (a) Slip velocity representation (b) RB convection in a cubic cavity.

Furthermore, a non-uniform grid is employed throughout the domain, with a finer grid resolution being used near the walls. Initial perturbations [6] are imposed to facilitate the transition of the flow from a laminar to a turbulent state. In the current investigation, an in-house off-lattice Boltzmann method solver [7, 8] is employed in the DNS mode for conducting simulations. The present implementation relies on a finite-difference-based framework for non-isothermal flows in the linear Oberbeck–Boussinesq regime. The LBM collision term is treated implicitly to improve the numerical stability of the scheme. Accurate characteristic-based treatment (of streaming) is obtained by a Lax–Wendroff scheme for time integration of the LBM equations. An explicit second-order central-difference scheme is used to discretize the advection terms. A sixth-order compact spatial filter is employed to remove oscillations caused by non-dispersive central difference schemes.

3 Results and Discussions

The simulations reveal significant differences in flow organization and heat transfer depending on the boundary conditions, as shown in **Figure 2** and **3**. For free-slip walls, a single large-scale vortex dominates the flow, enhancing coherent fluid motion and reducing flow reversal. In contrast, no-slip and finite-slip conditions produce smaller, chaotic vortices. This behavior arises from the fundamental distinction between the two conditions: free-slip walls permit tangential fluid motion along the boundary, whereas no-slip surfaces enforce zero fluid velocity at the boundary with a sharp velocity gradient. This high non-linearity in the no-slip case is the primary driver behind the formation of multiple eddy recirculation zones. Consequently, larger velocity fields are observed in the free-slip cases compared to the no-slip surface cases, resulting in a thinner thermal boundary layer thickness that significantly impacts the heat transfer rate. The free-slip condition on vertical walls significantly influences vortex structures, while free-slip on horizontal walls enhances velocity magnitudes and heat transfer.

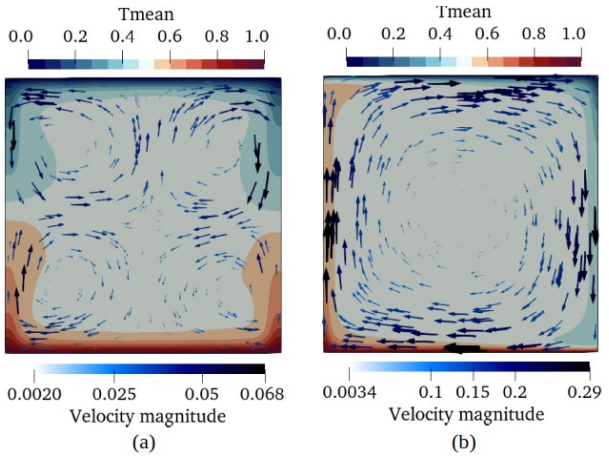


Figure 2 Mean velocity vectors and temperature distribution on the mid spanwise plane for temporal and spanwise averaged quantities. (a) No-slip (b) Free-slip.

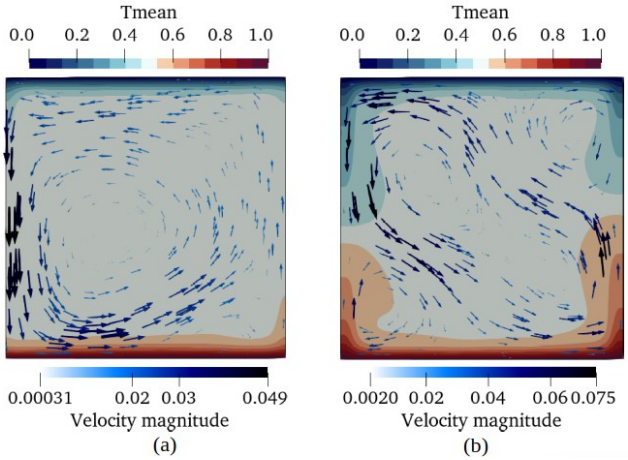


Figure 3 Mean velocity vectors and temperature distribution on the mid spanwise plane for temporal and spanwise averaged quantities. (a) Free-slip-Side (b) Slip-8.

Figures 4 and **5** show the horizontal (x-z-plane) averages of statistical quantities along the vertical axis for different boundary conditions at Ra = 2×10^6 and Pr = 4.38. The non-dimensionalized mean temperature profile along the cavity height is presented in **Figure 4(a)**. Under free-slip boundary conditions, the flow velocity is larger, leading to a thinner thermal boundary layer compared to the no-slip case, where a steeper gradient near the walls and uniformity in the central region is prominent. **Figure 4(b)** illustrates the non-dimensionalized vertical velocity component profile, where all cases, except FS and FS-side, exhibit a two-peak structure with a central dip, consistent with the (2, 2) vortex mode pattern in the flow as described above. By contrast, the free-slip cases FS and FS-side exhibits a parabolic profile variation, so that larger velocity values are present at the cavity center. Overall, the free-slip cases FS, FS-hori, and FS-side show larger velocity magnitudes than the no-slip case NS. Although the finite-slip cases deviate slightly from the no-slip case, they exhibit a similar overall profile. Finite-slip cases, positioned between no-slip and free-slip conditions, fall closer to no-slip behavior due to the use of small slip lengths.

DOI 10.5162/iCCC2026/P24

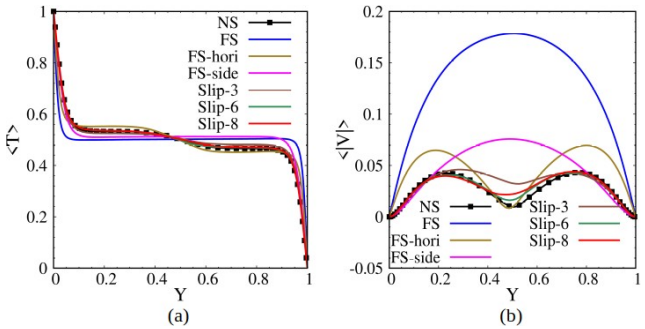


Figure 4 Normalized vertical profiles of horizontally averaged mean quantities. (a) Mean temperature (b) Mean vertical velocity magnitude.

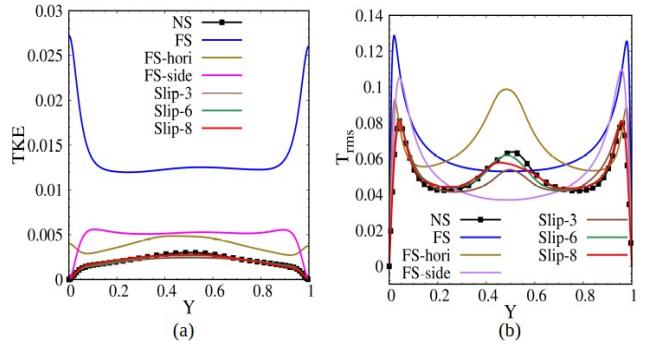


Figure 5 Normalized vertical profiles of horizontally averaged fluctuation quantities. (a) Turbulent kinetic energy (b) Temperature RMS fluctuations.

Figure 5(a) shows the distribution of turbulence intensity (turbulent kinetic energy) along the cavity height. The zero shear stress at the boundary in free-slip cases can lead to more intense turbulence, as the fluid near the walls experiences less resistance and can move more freely. Consequently, the turbulent kinetic energy (TKE) is generally higher for the FS case than the no-slip case. Although the slip cases show minor deviations from each other due to variations in slip length and anisotropy, their profiles largely resemble that of the no-slip case. Finally, **Figure 5(b)** illustrates the intensity of temperature fluctuations, quantified using the root mean square (RMS). The temperature RMS shows two peaks near the walls and a smaller peak at the center. However, in the case with a free-slip condition on the vertical sidewalls, the central peak disappears, and lower values are observed in the central region. The temperature RMS values are larger near the walls in free-slip due to more intense thermal mixing and less temperature fluctuations in the central region because the fluid near the boundaries is not as restricted by the no-slip condition, leading to a more uniform temperature distribution in the core. The temperature RMS variation in finite-slip cases is similar to that of the no-slip case.

The overall heat transfer rate for different wall conditions is quantified using the Nusselt number and reported in **Table 1**. The effects of different slip lengths and wall orientations, such as horizontal and vertical, are also reported at Ra = 2×10^6 and Pr = 4.38. Case names NS, FS, FS-side, FS-hori, and Slip 1-8 represent the following conditions: no-slip, free-slip on all walls, free-slip on vertical side walls, free-slip on horizontal walls, and finite-slip cases, respectively. The heat transfer rates are highest for free-slip conditions, with Nu increasing by up to 100% compared to no-slip cases. Because, in free-slip cases, the fluid experiences no frictional resistance at the walls, allowing for higher flow velocities near those boundaries. This results in a thinner thermal boundary layer, which significantly enhances the heat transfer rate. Finite-slip conditions with changing slip lengths also, however, show only marginal improvements over no-slip. The effect of surface condition on heat transfer rate for different fluids is presented in **Table 2**. The free-slip condition consistently yields higher Nusselt numbers compared to no-slip and finite-slip conditions across all Prandtl numbers investigated. Notably, the enhancement in heat transfer is most pronounced for Pr = 10, where the Nusselt number increases by 170% compared to the no-slip case. This is attributed to the combined effects of reduced viscous damping and enhanced thermal gradients at higher Prandtl numbers. Finite-slip boundary conditions result in heat transMean velocity vectors and temperature distri- bution on the mid spanwise plane for temporal and span- wise averaged quantities.fer rates that are nearly identical to those observed for no-slip conditions at all Prandtl numbers. This suggests that the finite-slip length considered here (l_s = 0.0005) is insufficient to significantly alter the flow dynamics or heat transfer characteristics compared to the no-slip condition. At the higher Rayleigh number Ra = 10^7 and Pr = 4.38, the heat transfer rates are significantly higher due to the increased buoyancy-driven turbulence. These results indicate that the finite-slip boundary conditions result in only minor changes in heat transfer at high Rayleigh numbers. A small increment in the Nusselt number is observed for the finite-slip case, l_s = 0.0001. Even at higher turbulent Rayleigh numbers, the effect of partial slip on heat transfer remains limited for the slip lengths considered in this study.

Case	Nu
NS	9.97
FS	19.99
FS-hori	15.97
FS-side	11.14
Slip-1	10.01
Slip-2	10.09
Slip-3	10.43
Slip-4	9.90
Slip-5	10.22
Slip-6	10.10
Slip-7	10.45
Slip-8	9.68

Table 1 Heat transfer in terms of Nusselt number Nu for various surface conditions at Ra = 2×10^6 and Pr = 4.38.

DOI 10.5162/iCCC2026/P24

Ra	Pr	Walls	Nu
2×10^6	0.786	No-slip	10.27
		Free-slip-all	21.43
		Slip l_s= 0.0005	10.30
	4.38	No-slip	9.97
		Free-slip-all	19.99
		Slip l_s= 0.0005	9.9
	10	No-slip	9.81
		Free-slip-all	26.55
		Slip l_s= 0.0005	9.75
10^7	4.38	No-slip	16.76
		Slip l_s= 0.0001	17.27
		Slip l_s= 0.0005	16.74

Table 2 Overall Nusselt number for various surface conditions and different fluids in terms of Pr and Ra.

Figure 6 shows how the flow patterns change when the wall condition is switched from a regular wall (no-slip) to a functionalized wall (free-slip). The comparison is made using the Lumley triangle analysis [9]. The anisotropy line is drawn along the central vertical axis. The influence of the walls on the turbulent states remains distinctly evident. In the no-slip case, a multitude of turbulent states is observed when moving from the bottom wall to the top wall, whereas a dominant single-component axisymmetric turbulent state is presented in the free-slip case. The comparison shows how surface boundary conditions fundamentally alter the nature of turbulence anisotropy, with free-slip walls promoting a simplified turbulent state relative to the more complex and spatially varying turbulence observed under no-slip conditions.

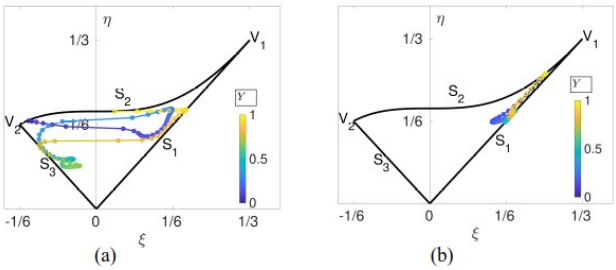


Figure 6 Lumley flow anisotropy analysis across the cavity height y (color) for the two velocity boundary conditions investigated. (a) No-slip (b) Free-slip.

4 Conclusions

In this study, we examined how different wall surface properties affect fluid motion and heat transfer in a heated and cooled system. We compared regular walls, where the fluid sticks to the surface, with coated walls, where the fluid can slide along the surface more easily. Using detailed computer simulations, we found clear differences in how the fluid moves and how heat is transported. Free-slip conditions resulted in an unsteady large-scale circulation pattern and enhanced heat transfer due to higher mean velocities and steeper temperature gradients near walls. In contrast, no-slip conditions exhibited diverse turbulent modes and flow reversals. The results highlight that wall boundary-slip can be a viable passive strategy to control thermal convection performance.

5 Acknowledgement

This research is supported by the German Federal Government, the Federal Ministry of Research, Technology and Space, and the State of Brandenburg within the framework of the joint project EIZ: Energy Innovation Center (project numbers 85056897 and 03SF0693A) with funds from the Structural Development Act (Strukturstärkungsgesetz) for coal-mining regions. Provision of computing resources by the National High Performance Computing Alliance (NHR-NORD@Göttingen) under the project number bbi00022 is also gratefully acknowledged.

6 References

[1] Choi, C.-H.; Kim, C.-J.: Physical Review Letters 96 (2006) H. 6, S. 066001.

[2] Mayeed, M. S.; Patnaik, S. S.; Mitchell, R.: In Proceedings of the ASME 2016 International Mechanical Engineering Congress and Exposition (IMECE2016). USA: ASME, 2016.

[3] Wu, C. H.; Huang, Y. S.; Kuo, L. S.; Chen, P. H.: International Journal of Heat and Mass Transfer 63 (2013) S. 249–254.

[4] Chillà, F.; Schumacher, J.: European Physical Journal E 35 (2012) H. 6, S. 58.

[5] Rothstein, J. P.: Annual Review of Fluid Mechanics 42 (2010) H. 1, S. 89–109.

[6] Schoppa, W.; Hussain, F.: Fluid Dynamics Research 26 (2000) H. 2, S. 119.

[7] Polasanapalli, S. R. G.; Anupindi, K.: Physics of Fluids 34 (2022) H. 3, S. 035125.

[8] Polasanapalli, S. R. G.; Anupindi, K.: International Journal of Heat and Mass Transfer 225 (2024) S. 125395.

[9] Lumley, J. L.: Advances in Applied Mechanics 18 (1979) S. 123–176.

DOI 10.5162/iCCC2026/P25

Mobile Shifted Excitation Raman Difference Spectroscopy Device for in vivo Skin Antioxidant Monitoring in Clinical Environments

Kay Sowoidnich[1], André Müller[1], Anne Pfennig[2], Svenja Karen Starke[2], Thomas König[2], Martin Schmidt-Hieber[2], Martin Maiwald[1]
[1]Ferdinand-Braun-Institut (FBH), Berlin, Germany, martin.maiwald@fbh-berlin.de
[2]Medizinische Universität Lausitz – Carl Thiem, Cottbus, Germany: a.pfennig@mul-ct.de; 2. Med. Clinic (Hematology, Oncology, Nephrology, Pneumology and Diabetology): m.schmidt_hieber@mul-ct.de

Abstract

We here present a novel mobile instrument based on shifted excitation Raman difference spectroscopy (SERDS) in combination with resonance Raman spectroscopy with the aim to analyze physiologically relevant carotenoid concentrations in human skin. Interferences from skin autofluorescence and ambient light are effectively addressed by our approach. We apply SERDS using an in-house developed micro-integrated 488 nm dual-wavelength diode laser system as excitation light source. Initial measurements on skin phantoms demonstrate the capability for quantitative carotenoid assessment with limits of detection of 0.05 nmol/g. Preliminary clinical data confirm that the SERDS system is highly effective for in vivo skin carotenoid assessment, making it a promising tool for non-invasive antioxidant monitoring.

1 Introduction

Assessing the status of antioxidants in human skin is of great importance for healthcare and nutritional applications [1]. An important group of antioxidants that can be found in the skin of humans are carotenoids. They are, however, only present at relatively low concentration (around 0.2-0.6 nmol/g) [2], thus complicating their detection. The current gold standard for carotenoid assessment in humans is based on drawing a blood sample that is subsequently analyzed by high-performance liquid chromatography (HPLC). While this approach is efficient in addressing the relevant concentration levels, it is invasive and requires time-consuming, complex, and expensive laboratory analyses [1].

Here, fast and non-destructive optical methods for non-invasive carotenoid detection directly in human skin show great promise as an alternative approach. Its ability to obtain molecule-specific information makes Raman spectroscopy a well-suited candidate for such a task. However, due to the weakness of the Raman effect, the technique suffers from dominant background interferences, that can originate, e.g., from skin autofluorescence or ambient lights. Moreover, the relevant carotenoid concentration range is not directly accessible using conventional Raman spectroscopy without any means of signal enhancement [3].

In this contribution, we combine shifted excitation Raman difference spectroscopy (SERDS) [4] with resonance Raman spectroscopy to generate an effective tool to address such challenges. SERDS uses a physical approach with two slightly shifted laser wavelengths that are used to record two Raman spectra. While Raman signals directly follow the applied wavelength shift, background interferences remain unaffected. Subtraction of the two recorded spectra thus provides a neat way of separating Raman signals from unwanted disturbing interferences.

The resonance Raman effect enables a selective enhancement of the Raman signals of carotenoids thus making accessible the relevant concentration ranges that are present in human skin.

For the application of the resonance Raman effect, an excitation wavelength that matches the absorption band of the target substances carotenoids is required. Moreover, for SERDS a laser excitation source with two distinct emission lines is necessary. To meet these requirements, our study applies an in-house-developed micro-integrated dual-wavelength diode laser system emitting at 487.6 nm and 487.9 nm [5].

2 Mobile SERDS instrument for in vivo skin investigations

The diode laser system with an emission around 488 nm is integrated into an in-house-realized fiber-coupled turn-key laser system as an operating environment providing the required laser injection currents as well as a Peltier-controlled heat sink. By means of an optical fiber, the laser radiation is then transferred to an internally designed and fabricated Raman probe head that is compatible for deployment in a clinical environment. Inside the probe head, the laser light is spectrally cleaned by means of a bandpass filter and subsequently directed through a sapphire window towards the skin.

To address the spatially heterogeneous distribution of the target substances (carotenoids) within the skin, a large excitation spot size of 3 mm diameter was chosen. The laser power at the sample position is limited to 14 mW resulting in a power density of 1981 W/m² that is less than the maximum permittable exposure defined by the standards DIN EN ISO 60825-1:2022-07 and ANSI Z136.1.

DOI 10.5162/iCCC2026/P25

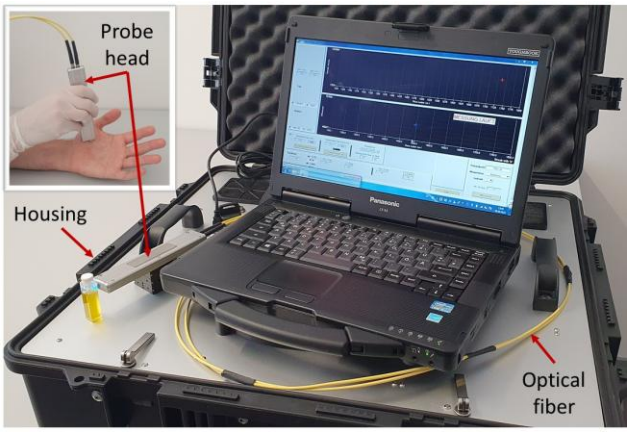


Fig. 1 Mobile SERDS system for the determination of the cutaneous carotenoid content. The instrument comprises a housing, a probe head, and a control notebook running the software for system operation and data management. The inset in the top left corner shows the probe head in contact with the skin during the measurement.

The scattered light emerging from the sample is collected in 180° backscattering geometry and reaches a Raman long-pass filter that only transmits the Stokes-shifted Raman light (located at longer wavelengths compared to the excitation radiation). The transmitted light is then launched into an optical fiber that transfers it towards a spectrometer (Tornado Hyperflux U1) with an attached CCD detector that is cooled down to an operation temperature of -10 °C.

Power supply of the turn-key laser system and the CCD detector is realized by means of a rechargeable $LiFePO_4$ battery that is built-in to the measurement system. A robust laptop computer (Panasonic Toughbook) serves for device control and data management. The battery capacity is sufficient for continuous operation of the system over a period of approximately 8 hours. A photograph of the device is depicted in Figure 1.

3 Measurement conditions

Recently, within the iCampµs II Research Network the mobile SERDS system was used in a clinical study at the Medical University Lausitz – Carl Thiem in Cottbus. The instrument was applied to the upper skin layers of patients with hematological neoplasia as well as to healthy control subjects in order to assess their cutaneous carotenoid levels. To determine reference serum carotenoid concentrations, HPLC of drawn blood samples from study participants was conducted at the German Institute of Human Nutrition (DIfE) in Potsdam-Rehbrücke.

As this was an 'other clinical investigation' according to the Medical Device Regulation (MDR, EU 2017/745) and the German Medical Device Law (MPDG), the study was approved by the Ethics Committee of the Brandenburg State Chamber of Physicians (Ethikkommission der Landesärztekammer Brandenburg) and registered with the German Medical Devices Information and Database System (Deutsches Medizinprodukte-Informations- und Datenbanksystem, DMIDS) of the Federal Institute for Drugs and Medical Devices (BfArM). The study was conducted according to the principles of the Declaration of Helsinki (1996) and Good Clinical Practice Guidelines. All subjects were informed about the study, provided written informed consent, and documented their agreement by signature.

4 Results and Discussion

A selected measurement on human skin is exemplarily presented in Figure 2. The Raman spectra excited at the two distinct excitation wavelengths were recorded with an exposure time for single spectra of 5 s and 10 accumulations were averaged each. The prominent Raman signal of carotenoids located around 1525 cm^{-1} is recognizable but it is superimposed on a strong signal background caused by skin autofluorescence.

The signal-to-background-noise ratio of both Raman spectra is 24. Application of SERDS efficiently separates the Raman signal from background interferences, leading to an improvement in the signal-to-background-noise ratio by approximately one order of magnitude.

Previously, the suitability for quantitative carotenoid assessment was demonstrated using skin phantoms with physiologically relevant beta-carotene concentrations. These investigations have shown that a limit of detection of 0.05 nmol/g of the target substance beta-carotene can be realized [3]. This value is 10-times below the mean beta-carotene concentration in skin thus qualifying the device for in vivo measurements. Subsequently, the instrument was successfully applied for in vivo studies investigating the skin carotenoid concentration with respect to nutritional effects [6] and consequences from anti-neoplastic treatment [7].

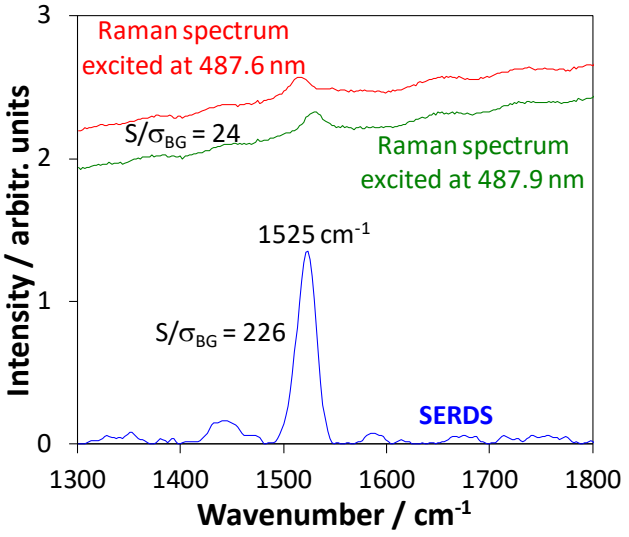


Fig. 2 Raman spectra (vertically shifted for clarity) of human skin excited at two slightly shifted wavelengths and the corresponding SERDS spectrum.

Initial evaluation of the data obtained during the study at the Medical University Lausitz – Carl Thiem indicates that there exists a significant correlation between carotenoid concentrations in blood (measured by HPLC) and skin (obtained using SERDS). Furthermore, carotenoid levels in healthy controls are clearly elevated compared to patients and a positive influence of plant-based diets could be observed. An in-depth analysis of the large data set acquired during the study is currently in progress.

5 Conclusions

Our study demonstrates the capability of the mobile SERDS instrument for cutaneous in vivo carotenoid detection. The device thus paves the way for quantitative non-invasive carotenoid measurements in the frame of future clinical studies.

6 Acknowledgement

This work was supported by the Federal Ministry of Research, Technology and Space (BMFTR) within the projects iCampus (Grant numbers 16ES1132 and 16ME0425) and Research Fab Microelectronics Germany – FMD (Grant number 16FMD02).

Additional thanks go to Madlen Löbel and her team at the Clinical Trial Unit of the Medical University Lausitz – Carl Thiem, whose support in data acquisition and patient/participant management was highly appreciated, and to Daniela Weber and her team at the German Institute of Human Nutrition Potsdam-Rehbrücke (DIfE) for conducting the HPLC analyses of blood sera.

7 References

[1] Obana, A.: Measurement of skin carotenoids and their association with diseases: a narrative review. BBA - Molecular and Cell Biology of Lipids (2025) 1870, 159612.

[2] Peng, Y.M.; Peng, Y.S.; Lin, Y.: A nonsaponification method for the determination of carotenoids, retinoids, and tocopherols in solid human tissues. Cancer Epidemiol. Biomarkers Prev. (1993) 2, 139-144.

[3] Sumpf, B.; Braune, M.; Maiwald, M.; Darvin, M.E.; Lademann, J.; Tränkle, G.: A diode laser based clinical diagnostic system using shifted excitation resonance Raman difference spectroscopy for the in vivo detection of ß-carotene in human skin. Proc. SPIE (2020) 11257, 1125707.

[4] Shreve, A.P.; Cherepy, N.J.; Mathies, R.A.: Effective Rejection of Fluorescence Interference in Raman Spectroscopy Using a Shifted Excitation Difference Technique. Appl. Spectrosc. (1992) 46, 707-711.

[5] Braune, M.; Maiwald, M.; Sumpf, B.; Tränkle, G.: 2 nm continuously tunable 488 nm micro-integrated diode laser based SHG light source for Raman spectroscopy. Proc. SPIE (2016) 9731, 973116.

[6] Henning, T.; Wagner, P.; Gedat, E.; Kochlik, B.; Kusch, P.; Sowoidnich, K.; Vastag, M.; Gleim, J.; Braune, M.; Maiwald, M.; Sumpf, B.; Grune, T.; Weber, D.: Evaluation of Modern Approaches for the Assessment of Dietary Carotenoids as Markers for Fruit and Vegetable Consumption. Nutrients (2023) 15, 1665.

[7] Jung, S.; Darvin, M. E.; Schleusener, J.; Thiede, G.; Lademann, J.; Braune, M.; Maiwald, M.; Sumpf, B.; Tränkle, G.; Kutzer, D.; Elban, F.; Fuss, H.: In vivo detection of changes in cutaneous carotenoids after chemotherapy using shifted excitation resonance Raman difference and fluorescence spectroscopy. Skin Res. Technol. (2020) 2, 301-307.

DOI 10.5162/iCCC2026/P26

Digital unterstützte Betreuung von Pflegeeinrichtungen – Vorstellung des DUAB-Projektes

Digitally supported care for nursing homes – Presentation of the DUAB project

Dr. med. Kay-Patrick Braun, Regina Märker, Amadeus-Valentin Klement, MVZ Dr. Braun GmbH, Cottbus, Deutschland
kay-p.braun@web.de

Kurzfassung

Der Wandel der Altersstruktur hin zu einer zunehmenden Alterung hat eine vermehrte Versorgung von Patient:innen in Pflegeeinrichtungen zur Folge. Dies zieht einen erhöhten hausärztlichen Versorgungsbedarf in diesen Wohnstätten nach sich. Dem gegenüber steht ein zunehmender Mangel an hausärztlichen Kolleg:innen. Satt des bisher üblichen, zeitaufwendigen Hausbesuches sind neue Versorgungsformen von Nöten. Das DUAB-Projekt sieht eine effektive Betreuung von Patient:innen in Pflegeeinrichtungen in Akutsituationen unter Zuhilfenahme digitaler Möglichkeiten zur raschen Entscheidungsfindung vor, ohne dass ein zeitaufwendiger persönlicher Besuch durch die Hausärztin oder den Hausarzt notwendig ist. Die Datenerhebung und Auswertung erfolgt unter wissenschaftlichen Gesichtspunkten und Fragestellungen.

Abstract

The shift in the age structure towards an increasingly ageing population is leading to more patients being cared for in nursing homes. This results in an increased need for GP care in these residential facilities. This is offset by a growing shortage of GPs. Instead of the time-consuming home visits that have been the norm up to now, new forms of care are needed. The DUAB project aims to provide effective care for patients in nursing homes in acute situations with the help of digital tools for rapid decision-making, without the need for time-consuming personal visits by GPs. Data collection and evaluation are carried out according to scientific criteria and questions.

1 Hintergründe

Die Altersstruktur der Bevölkerung in Deutschland unterliegt einem Wandel hin zu einer zunehmenden Alterung. Im Jahre 2020 betrug der Anteil von Menschen über 65 Jahre ca. 22%. Für 2040 ist ein Anstieg auf 28% prognostiziert. Die Ursache liegt in einer zunehmenden Lebenserwartung. Hier ist ein kontinuierlicher Anstieg zu verzeichnen. Sie liegt aktuell für Frauen bei 81 Jahren und für Männer bei 76 Jahren. Dadurch bedingt steigt auch der Bedarf an Pflegeeinrichtungen. Es werden Zahlen berichtet, dass 2019 rund 818.000 Menschen in Pflegeeinrichtungen lebten. Dies stellt einen Anstieg um 33% im Vergleich zu 1999 dar. Ein weiterer Anstieg ist zu erwarten.

Die steigende Lebenserwartung verbunden mit vermehrtem Auftreten von Krankheiten des Alters sowie die Zunahme der Zahl von in Pflegeeinrichtungen lebenden Menschen hat einen erhöhten Bedarf an hausärztlicher Versorgung zur Folge. Jedoch ist statt der dringend Benötigten steigenden Zahl an Hausärzt:innen ein Rückgang zu verzeichnen. Waren es im Jahre 1995 noch ca. 56.000, so wird für das Jahr 2024 eine Zahl von ca. 45.000 hausärztlich tätigen Kolleg:innen angegeben. Wichtig ist hierbei zu berücksichtigen, dass im Jahre 1996 der Anteil der über 65-jährigen Kolleg:innen bei 4,9% lag und dieser 2024 auf 20,3% angestiegen ist.

Eine mögliche Ursache für diesen Anstieg liegt darin, dass zum 01.01.2009 die Altersbegrenzung für niedergelassene Ärzt:innen weggefallen ist. Bis dahin galt, dass mit Vollendung des 68. Lebensjahres automatisch die Zulassung zur vertragsärztlichen Versorgung erlischt.

Weitere wichtige Aspekte in der Versorgungsstruktur stellen die Zulassungsmöglichkeiten dar. So wurde im Jahre 2004 die Gründung von Medizinischen Versorgungszentren (MVZ) ermöglicht. Zunächst waren hierbei mehrere Fachrichtungen notwendig. Seit 2012 sind nun auch fachgleiche MVZ möglich. Einen weiteren wichtigen Meilenstein bildet das Vertragsarztrechtsänderungsgesetz aus dem Jahre 2007. Neben dem bereits erwähnten Wegfall der Altersbegrenzung beinhaltet es eine deutliche Erweiterung der Gestaltung der Versorgung durch Schaffung von deutlichen umfangreicheren Möglichkeiten zur Tätigkeit in Anstellung.

Wie angespannt die Situation aktuell ist, unterstreicht einmal mehr, dass im Gebiet der Kassenärztlichen Vereinigung des Landes Brandenburg aktuell in 21

DOI 10.5162/iCCC2026/P26

Regionen, in denen aufgrund einer drohenden Unterversorgung Zulassungen im hausärztlichen Bereich gefördert werden. Für das Stadtgebiet Cottbus werden aktuell Zulassungen und Anstellungen finanziell unterstützt.

2 Ist-Zustand

Aktuell erfolgt die Versorgung der Patient:innen in Pflegeeinrichtungen bei fehlender Mobilität durch die betreuenden Hausärzt:innen im Hausbesuch. Die Kommunikation zwischen der Einrichtung und der Praxis wird häufig per Fax oder telefonisch realisiert. In dringenden Fällen wird bei Nichterreichbarkeit der Hausärztin bzw. des Hausarztes häufig der Rettungsdienst in Anspruch genommen.

In den meisten Fällen haben Pflegeeinrichtungen und Arztpraxen jeweils ihre eigene Dokumentation. Ansätze für eine gemeinsame Dokumentation bestehen, sind aber noch die Ausnahme.

Auf hausärztlicher Seite gibt es für die Unterstützung zur Versorgung in Pflegeeinrichtungen spezialisierte Ausbildungen für die medizinischen Fachangestellt:innen (MFA) wie AGNES oder NÄPA. Diese haben umfangreichere Kompetenzen in der Betreuung. Allerdings gibt es regional nicht nur Probleme in der ärztlichen Versorgung sondern auch in der Gewinnung von Fachpersonal.

Darüber hinaus besteht teilweise auch bereits ein Problem für zugezogene Bewohner:innen in Pflegeeinrichtungen eine Hausärztin oder einen Hausarzt zu finden.

3 Verbesserung der Versorgungssituation

Ist eine Verschlechterung des Allgemeinzustandes von Patient:innen in Pflegeeinrichtungen zu verzeichnen, die nicht mobil sind, wird häufig ein Hausbesuch durch die Hausärztin bzw. den Hausarzt angefordert. Zur Einschätzung des aktuellen Zustandes und Festlegung des weiteren Procedere ist dies durchaus sinnvoll. Allerdings sind hierfür einige Voraussetzung notwendig:

1. Die betreffende Patient:in muss über eine Hausärztin bzw. einen Hausarzt verfügen.

2. Die betreuende Hausärzt:in muss erreichbar sein.

Darüber hinaus ergeben sich aus dieser Situation folgende Herausforderungen und Probleme:

1. In der Regel ist kein sofortiger Besuch aus der Sprechstunde heraus möglich. Eine Entscheidung erfolgt erst mit zeitlicher Verzögerung.

2. Der Hausbesuch stellt einen erheblichen Zeitaufwand für die Hausärztin/ den Hausarzt dar.

3. Wird die Hausärztin oder der Hausarzt nicht erreicht, ist ein Hausbesuch nicht möglich oder besteht keine hausärztliche Betreuung wird häufig der Rettungsdienst alarmiert.

Bei einem Hausbesuch durch die Hausärztin bzw. den Hausarzt erfolgt in der Regel vor Ort eine Anamneseerhebung, eine klinische Beurteilung des Allgemeinzustandes, eine körperliche Untersuchung mittels „sehen, fühlen, hören“. Zusätzlich ist noch die Bestimmung der Sauerstoffsättigung möglich. Teilweise erfolgt die Untersuchung mittels eines mobilen EKG´s.

Bei kritischer Betrachtung muss man erkennen, dass für die überwiegenden Gegenstände des Hausbesuches keine persönliche Präsenz der Ärztin oder des Arztes vor Ort nötig wäre. Die Anamneseerhebung ist problemlos auf die Entfernung in identischer Qualität möglich. Eine visuelle Beurteilung wäre mittels digitaler Möglichkeiten mit vergleichbarem Ergebnis zu erreichen. Hierbei ist der Einsatz einer Videobrille denkbar. Die Messung der O2-Sättigung kann problemlos vor Ort erfolgen. Und auch die Ableitung eines EKG´s sowie die Auskultation kann mit bereits am Markt befindlichen digitalen Lösungen in vergleichbarer Qualität auf die Entfernung erfolgen.

DOI 10.5162/iCCC2026/P26

3.1 Notwendige Hilfsmittel

Um eine digital basierte Betreuung von Patient:innen in Pflegeeinrichtungen zu ermöglichen, sind einige Voraussetzungen zu schaffen. Diese sind in der Regel jedoch nicht sonderlich aufwendig und bereits überwiegend verfügbar.

Grundvoraussetzung ist eine ausgebildete Mitarbeiter:in vor Ort. Dies kann zum einen eine nichtärztliche Praxisassistentin der betreuenden Hausärzt:in sein. Zum anderen kann diese Funktion ebenfalls durch eine speziell geschulte Mitarbeiter:in der Pflegeeinrichtung erfolgen.

Darüber hinaus ist eine Videobrille erforderlich. Die nichtärztliche Praxisassistent:in beziehungsweise geschulte Mitarbeiter:in der Pflegeeinrichtung trägt diese vor Ort, um den Kontakt zu der Patientin beziehungsweise dem Patienten herzustellen.

Auf der anderen Seite ist für die betreuende Ärztin beziehungsweise den betreuenden Arzt ein entsprechender PC mit der Möglichkeit zur Videosprechstunde erforderlich.

Einen weiteren wichtigen Aspekt stellt eine Dokumentation der Pflegeeinrichtung dar, auf die die Ärztin beziehungsweise der Arzt direkt Zugriff hat. Hierin sind Vitalwerte, wie Gewicht, Puls, Blutdruck und ggf. Blutzucker sowie Ausscheidung zur Beurteilung direkt ersichtlich. Darüber hinaus können die aktuelle Medikation, ggf. verabreichte Insulindosen oder Bedarfsmedikationen und eventuelle Einträge anderer Ärzt:innen eingesehen werden.

Des Weiteren erscheint ein digitales Stethoskop mit Möglichkeit zur Übertragung, welches so am Markt bereits verfügbar ist, sinnvoll. Dieses kann die geschulte Mitarbeiter:in vor Ort an den entsprechenden Untersuchungsstellen unter ständiger videobasierter visueller Kontrolle der Ärztin bzw. des Arztes aufsetzen. Somit ist eine Untersuchung von Herz, Lungen und Bauch möglich.

Zur weiteren klinischen Einschätzung der Situation kann darüber hinaus ein Pulsoxymeter zum Einsatz kommen. Hiermit sind Herzfrequenz und Sauerstoffsättigung zu beurteilen. Auch ein mobiles EKG, was durch die Mitarbeiter:in vor Ort aufgesetzt und digital an die Ärzt:in übertragen wird, kann zur Beurteilung der Situation beitragen.

Neben den genannten personellen und technischen Voraussetzungen sind noch weitere Aspekte zu bedenken. Hierbei ist eine stabile Internetverbindung zu erwähnen, was natürlich die Basis für diese Form der Diagnostik und Behandlung darstellt. Wenn es sich um eine mobilphone basierte Variante handelt, können sich teils durch bauliche Einschränkungen mit Stahlbeton bedingten Abschirmungen Probleme ergeben. Darüber hinaus bestehen insbesondere in ländlichen Regionen sowohl bei mobilem als auch festnetzbasiertem Internet Grenzen in der Netzabdeckung.

Einen weiteren wesentlichen, noch zu klärenden Aspekt stellt die Umsetzung des Datenschutzes dar. Zum einen ist natürlich eine Einwilligung der Patient:in erforderlich. Hierbei sind bei Vorliegen einer beispielsweise demenzbedingten Einschränkung der persönlichen Einwilligungsfähigkeit eventuell bestehende Betreuungsverhältnisse zu beachten. Darüber hinaus stellt der Datenschutz ebenfalls bei den räumlichen Gegebenheiten einer videobasierten Konsultation einen ganz zentralen Aspekt dar. Die Diskretion und Schweigepflicht sind hierbei

selbstverständlich uneingeschränkt, wie bei einer persönlichen Betreuung zu beachten.

3.2 Anwendungsbeispiele

Der Einsatz einer digitalen Sprechstunde erscheint neben der routinemäßigen Konsultation insbesondere für Notfallsituation sinnvoll. In diesen Fällen kann sowohl für eigene Patient:innen, aber auch denkbar für andere eine kurzfristige Beurteilung der aktuellen Situation mit Entscheidung über das weitere Procedere erfolgen. Dies kann in abwartendem Offenlassen, der Veranlassung einer stationären Einweisung, bei Fremdpatient:innen der Rücksprache mit der betreuenden Hausärztin bzw. dem betreuenden Hausarzt oder der Verordnung einer Therapie bestehen.

Denkbar wäre auch die Einrichtung einer zentralen digitalen Notfallsprechstunde für mehrere Einrichtungen.

Durch den Einsatz einer videobasierten Konsultation mit den genannten Vorteilen kann zum einen die Zufriedenheit von Bewohner:innen, Angehörigen und auch der Pflegekräfte erreicht werden, zum anderen lassen sich unnötige Notfalleinsätze des Rettungsdienstes und nicht indizierte Einweisungen reduzieren.

Im Rahmen eines Gesamtkonzeptes wäre darüber hinaus der Einsatz einer Praxisassistentin (PA) mit spezialisierter Ausbildung als erste Ansprechpartnerin und für die Triagierung denkbar.

3.3 Zeitlicher Ablauf

In der Pilotphase ist zunächst zum Aufbau der strukturellen Gegebenheiten sowie Testung und Optimierung der Abläufe die Zusammenarbeit zwischen der Praxis der MVZ Dr. Braun GmbH und einer regionalen Pflegeeinrichtung geplant.

Im weiteren Verlauf sollen zunächst weitere betreute Pflegeeinrichtungen eingeschlossen werden.

Nach Abschluss dieser Pilotphase und Schaffung stabiler Strukturen ist die Weitergabe des Konzeptes an interessierte Praxen regional als auch überregional vorgesehen.

3.4 Wissenschaftliche Fragestellungen

Aufgrund der Relevanz der Thematik ist auch eine wissenschaftliche Begleitung des Projektes mit entsprechender Datenerhebung, Auswertung und Publikation Bestandteil des Projektes.

Durch eine Evaluierung der Ergebnisse mit noch zu entwickelnden und zu testenden Fragebögen ist die Beantwortung nachfolgend genannter Fragestellungen denkbar:

1. Vergleich der Zufriedenheit von Bewohner:innen, Angehörigen und Mitarbeiter:innen der Pflegeeinrichtung von im Rahmen dieses Projektes betreuten Wohnstätten mit herkömmlichen?

2. Zahl der stationären Einweisungen von DUAB im Vergleich zu Vergleichsgruppen?

3. Als langfristige Fragestellung: entsteht ein Lebenszeitgewinn durch DUAB?

Für die wissenschaftliche Betreuung sind neben der Entwicklung der Fragebögen eine detaillierte Projektbeschreibung, ein Antrag bei der Ethikkommission sowie eine Anmedlung bei dem Deutschen Register klinischer Studien (DRKS) erforderlich.

3.5 Finanzierung

Ein Antrag bei der Kassenärztlichen Vereinigung Brandenburg (KVBB) auf Gewährung einer Förderung gemäß der Maßnahme „Förderung telemedizinischer Versorgungsformen sowie telemedizinischer Kooperationen gem. §1 Abs. 1 Punkt IX. und XI. der Richtlinien zur Verwendung von Mitteln des Strukturfonds gem. §105 SGB V der KVBB“ ist bereits gestellt.

4 Literatur

[1] Statistisches Bundesamt.
[2] Bertelsmann Stiftung.
[3] Deutsches Institut für Wirtschaftsforschung (DIW).
[4] Kassenärztliche Bundesvereinigung (KBV).
[5] Kassenärztliche Vereinigung Brandenburg (KVBB).

DOI 10.5162/iCCC2026/P27

Mikrointegriertes Quantenlichtmodul für die hyperspektrale Bildgebung im mittleren Infrarot in der Krebsdiagnostik

Nils Werner[1], Susann-Alice Seeger[1], Felix Mauerhoff[1], Atta Ur Rehman Sherwani[2], Alexander Sahm[1], Johannes Zender[1], David Feise[1], Gunnar Blume[1], Sven Ramelow[1], Katrin Paschke[1]
[1]Ferdinand-Braun-Institut (FBH), Berlin, Deutschland
[2]Institut für Physik, Humboldt-Universität, Berlin, Deutschland
nils.werner@fbh-berlin.de

Kurzfassung

Die spektrale Analyse von Biomolekülen im mittleren Infrarot (MIR) Bereich bietet ein enormes Potential für die Krebsdiagnostik, da sich bio-chemische Veränderungen Label-frei nachweisen lassen. Durch Anwendung von Quanteneffekten mit dem Konzept „Sensing mit nichtdetektiertem Licht“ lässt sich die Spektroskopie im MIR durchführen bei ausschließlicher Detektion im nahen Infrarot und dabei zahlreiche Herausforderungen klassischer MIR-Quellen und Detektoren umgehen. Mit der Entwicklung einer maßgeschneiderten Diodenlaserquelle bei 720 nm und der Mikrointegration wesentlicher Bestandteile des oft komplexen interferometrischen Laboraufbaus steht nun ein kompaktes Quantenlichtmodul zur Verfügung, das das Konzept einer breiteren Anwendung näherbringt. Das Messprinzip konnte zwischen Wellenzahlen von 2700 cm^{-1} bis 3100 cm^{-1} demonstriert werden. Kombiniert mit klassischer Mikroskopie zu hyperspektraler Bildgebung besteht so die Möglichkeit die Krebsdiagnostik entscheidend zu verbessern.

Abstract

Spectral analysis of biomolecules in the mid-infrared (MIR) range offers enormous potential for cancer diagnostics, as it allows to detect biochemical changes label-free. By applying quantum effects in using the concept of sensing with undetected light, spectroscopy can be performed in the MIR while the detection is solely in the near-infrared, thereby overcoming several challenges associated with conventional MIR sources and detectors. With the development of a customized diode laser source at 720 nm and the micro-integration of essential components of the often complex interferometric laboratory setup, a compact quantum light module is now available that brings the concept closer to broader application. The measurement principle could be demonstrated between wave numbers of 2700 cm^{-1} to 3100 cm^{-1}. Combined with classical microscopy for hyperspectral imaging, cancer diagnostics could be significantly improved.

1 Einleitung

Der Spektralbereich im mittleren Infrarot (MIR) ist von besonderer Bedeutung für verschiedenste Analysemethoden in Chemie, Biologie oder Medizin. In diesem Spektralbereich lassen sich über die Molekülschwingungen diverse Biomoleküle und ihre Bindungseigenschaften charakterisieren. Die so mögliche molekulare Analyse bietet ein enormes Potential in der Krebsdiagnostik [1, 2], indem biochemische Veränderungen direkt im Gewebe Label-frei detektiert werden können.

Trotz dieses Potenzials sind die derzeit verfügbaren Lichtquellen und Detektoren im MIR-Bereich mit erheblichen Herausforderungen konfrontiert. Die Systeme sind oft kostenintensiv, wenig effizient oder decken nicht den gesamten geforderten spektralen Bereich ab. Daraus ergeben sich in der Anwendung diverse Nachteile wie hohe Lichtintensitäten auf der Probe, lange Messzeiten oder ein technisch aufwändiger Betrieb.

Um die Limitierungen bei der Erzeugung und Detektion von Strahlung im MIR zu umgehen, können Methoden der Quantenoptik verwendet werden. Die spontane parametrische Fluoreszenz (SPDC) ermöglicht die Erzeugung von breitbandigem Licht im mittleren Infrarot mittels Laserquellen im sichtbaren beziehungsweise nahinfraroten (NIR) Spektralbereich. Über nichtlineare Interferometer lässt sich schließlich das Konzept Sensing mit nichtdetektiertem Licht realisieren, was ausschließlich mit Detektoren im NIR auskommt, eine Probe aber im MIR analysiert wird. Im Labor konnte dieser Ansatz bereits für die Spektroskopie im MIR genutzt werden [3]. Da die Probe nur geringen Lichtintensitäten ausgesetzt wird und nur Strahlung im NIR mit etablierter Technik detektiert wird, ergibt sich durch diese Methode ein enormes Potential für die Biomedizin und Krebsdiagnostik. Damit diese auch Anwendung im klinischen Umfeld finden, ist es notwendig, das komplexe interferometrische Laborsetup in robusten und kompakten Modulen zu integrieren.

In dieser Arbeit wird eine Lichtquelle im MIR vorgestellt, bei der wesentliche Baugruppen in einem Mikromodul integriert sind. Dabei wird das Konzept Sensing mit nichtdetektiertem Licht in einem Wellenzahlenbereich von 2700 cm^{-1} bis 3100 cm^{-1} demonstriert.

2 Sensing mit nichtdetektiertem Licht

Das Konzept „Sensing mit nichtdetektiertem Licht“ kann über eine Interferometer-Anordnung mit einem Kristall mit nichtlinearen optischen Eigenschaften realisiert werden.

DOI 10.5162/iCCC2026/P27

Der in dieser Arbeit genutzte Aufbau ist schematisch in **Bild 1** dargestellt.

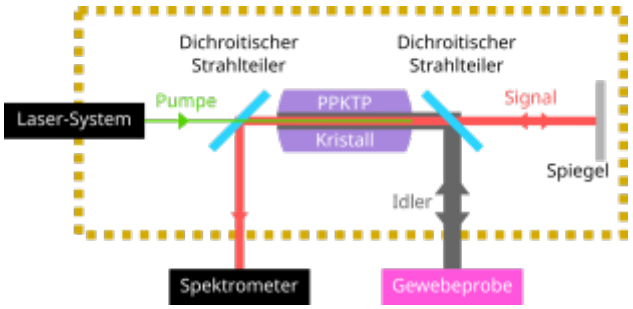


Bild 1 Schema des Sensorprinzips mit nichtdetektiertem Licht. Der gestrichelte Kasten markiert die Baugruppen, die als Quantenlichtmodul mikrointegriert wurden.

Das Laser-System stellt Laserstrahlung bei 720 nm bereit, die als Pumpe bezeichnet wird. Die Laserstrahlung passiert einen dichroitischen Strahlteiler und wird in einen periodisch gepolten Kaliumtitanylphosphat-Kristall (PPKTP) eingekoppelt. Die Front- und Rückfacette des Kristalls ist jeweils mit einer sphärischen Krümmung sowie einer Verspiegelung für die Wellenlänge der Pumpe versehen, wodurch eine Resonanz-Kavität für diese gebildet wird. Sofern die Kristalltemperatur und die Pumpen-Wellenlänge präzise aufeinander abgestimmt sind, wird die Pumpe resonant überhöht. Im Kristall wird über den SPDC-Prozess die sogenannte Signal- und Idler-Strahlung generiert.
Die periodische Polung des Kristalls ist derart gewählt, dass die Signal-Strahlung eine Wellenlänge um 910 nm und die Idler-Strahlung eine Wellenlänge um 3400 nm beziehungsweise eine Wellenzahl um 2900 cm^{-1} aufweist. Besonderes Merkmal dabei ist, dass Signal- und Idler-Strahlung jeweils spektral breitbandig sind [4] und so gerade im MIR mehrere hundert Wellenzahlen gleichzeitig abgedeckt werden können.
Signal und Idler werden aus dem PPKTP-Kristall in Vorwärts- und Rückwärtsrichtung emittiert. In Vorwärtsrichtung trennt ein dichroitischer Strahlteiler Signal und Idler. Die Idler-Strahlung wird für die Untersuchung von z.B. einer Gewebeprobe verwendet, wobei wichtig ist, dass die Strahlung von der Probe selbst oder einem dahinter liegenden Spiegel in sich zurückreflektiert wird. Die Signal-Strahlung wird ebenfalls von einem Spiegel in sich zurückreflektiert, dessen Abstand zum Kristall gleich dem Probenabstand zum Kristall gewählt ist. Sowohl die Signal- als auch die Idler-Strahlung passiert den Kristall erneut und überlagert sich so mit der in Rückwärtsrichtung erzeugten Strahlung. Bei der Überlagerung kann es zur Quanteninterferenz kommen, abhängig vom optischen Weg (und Interaktion mit der Probe) der Idler-Strahlung. Aufgrund des SPDC-Prozesses sind Signal und Idler quantenmechanisch verschränkt und bilden einen Zweiphotonenzustand. Wird die Idler-Strahlung von der Probe vollständig reflektiert und hat exakt denselben optischen Weg wie die Signal-Strahlung zu Spiegel, so ist die Quanteninterferenz maximal. Wird die Idler-Strahlung jedoch durch die Probe abgeschwächt, so reduziert sich die Stärke der Interferenz. Sofern die Signal- und Idler-Strahlung spektral breitbandig ist, lässt sich somit anhand der Stärke des Interferenzsignals eine Aussage über Absorptionseigenschaften der Probe im MIR erhalten. Aufgrund des Zweiphotonenzustandes kann die Interferenz sowohl mit der Idler- als auch mit der Signal-Strahlung (die niemals mit der Probe interagiert hat) gemessen werden. Der dichroitische Strahlteiler vor dem Kristall ist daher so ausgeführt, dass nur die Signal-Strahlung reflektiert und zu einem Spektrometer geführt wird. Dort erfolgt die Messung im NIR, während Messinformationen der Probe im MIR erhalten werden.
Um ein möglichst hohes Signal-Rausch-Verhältnis bei der Messung zu erzielen, ist eine effiziente Erzeugung der Signal- und Idler-Strahlung mittels SPDC erforderlich. Dafür ist zum einen eine hohe optische Leistung der Pumpe im Wattbereich erforderlich, zum anderen sorgt die resonante Überhöhung im Kristall bauartbedingt für eine hohe Erzeugungsrate der Signal- und Idler-Strahlung. Auch wenn für die Erzeugung der MIR-Strahlung hohe optische Leistungen benötigt werden, so ist die optische Leistung der Idler-Strahlung, die mit der Probe wechselwirkt kleiner als 1 µW.
Zur Realisierung des Konzeptes ist es sinnvoll, vor allem den interferometrischen Aufbau um den Kristall herum in ein kompaktes Modul zu integrieren. In **Bild 1** sind diese wesentlichen Baugruppen durch den gestrichelten Kasten markiert.

3 MIR-Quantenlichtmodul

Das mikrointegrierte Sensing-Modul umfasst Teile des Laser-Systems bei 720 nm sowie den interferometrischen Aufbau um den PPKTP-Kristall. Im Folgenden wird auf diese Baugruppen näher eingegangen.

3.1 Diodenlaserquelle bei 720 nm

Das Lasersystem ist als Hochleistungs-Diodenlaserquelle basierend auf AlGaAs und dem Master-Oscillator-Power-Amplifier (MOPA) Konzept ausgeführt. Die vertikale Halbleiterstruktur wurde speziell entwickelt und ist in Referenz [5] näher beschrieben. Der schematische Aufbau des MOPA-Konzeptes ist in **Bild 2** gezeigt.
Ein Rippenwellenleiterlaser mit integriertem Bragg-Gitter (DBR-RWL) fungiert als Master Oscillator und emittiert Laserstrahlung bei 720 nm. Die Laserstrahlung ist durch eine geringe spektrale Linienbreite in der Größenordnung von 1 MHz und eine nahezu beugungsbegrenzte Strahlqualität gekennzeichnet. Die Ausgangsleistung ist jedoch durch die hohe Leistungsdichte im Rippenwellenleiter auf unter 100 mW beschränkt. Um dennoch die benötige Leistung im Watt-Bereich zu erreichen, wird die Laserstrahlung in einem Trapezverstärker (TPA) als Power Amplifier nachverstärkt. Durch trapezförmige Aufweitung in einer Strahldimension können so mehr als 3 W optische Leistung erzielt werden, ohne dass kritische Leistungsdichten auftreten.

DOI 10.5162/iCCC2026/P27

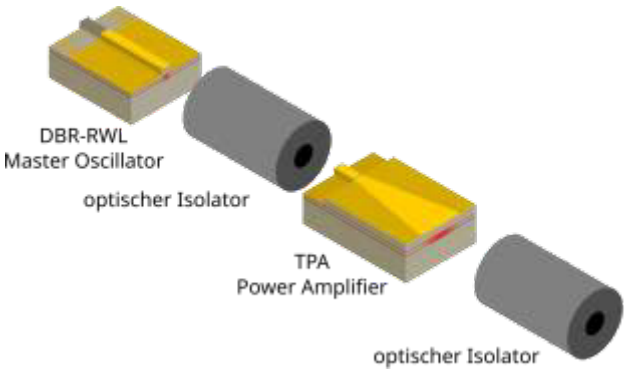


Bild 2 Schematische Darstellung des verwendeten Laser-Systems.

In **Bild 3** sind die elektro-optischen und spektralen Emissionseigenschaften des MOPAs gezeigt.

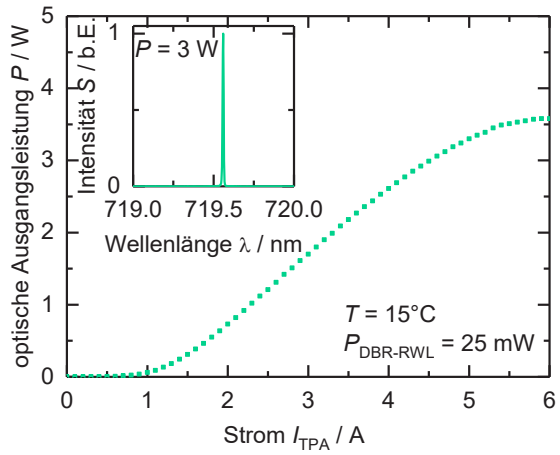


Bild 3 Optische Ausgangleistung des verwendeten Laser-Systems als Funktion des Injektionsstroms des TPAs. Das Inset zeigt das Emissionsspektrum bei einem Arbeitspunkt mit 3 W Ausgangsleistung.

Der DBR-RWL wird bei einer konstanten Ausgangsleistung von 25 mW betrieben, während der Injektionsstrom des TPAs variiert wird. Ab etwa 1 A nimmt die Ausgangsleistung nahezu linear zu und beginnt ab etwa 5 A, bei über 3 W optischer Leistung, aufgrund thermischer Effekte zu stagnieren. Die Emissionswellenlänge bleibt dabei unverändert und ist durch den DBR-RWL bestimmt. Das Inset in **Bild 3** zeigt exemplarisch das Emissionsspektrum bei 3 W Ausgangsleistung. Da optische Rückreflexe zum DBR-RWL die Emissionswellenlänge verändern können, müssen diese bestmöglich unterdrückt werden. Sowohl der TPA als auch der interferometrische Aufbau um den Kristall können Rückreflexe erzeugen, weshalb optische Isolatoren sowohl vor als auch hinter dem TPA platziert werden müssen.

3.2 Mikrooptischer Aufbau

Der interferometrische Aufbau wird mitsamt dem Kristall sowie Teilen des Laser-Systems auf möglichst kleinem Raum mikrointegriert. In **Bild 4** ist die Aufsicht des Aufbaus, der für die Integration auf einer mikro-optischen Bank vorgesehen ist, gezeigt.

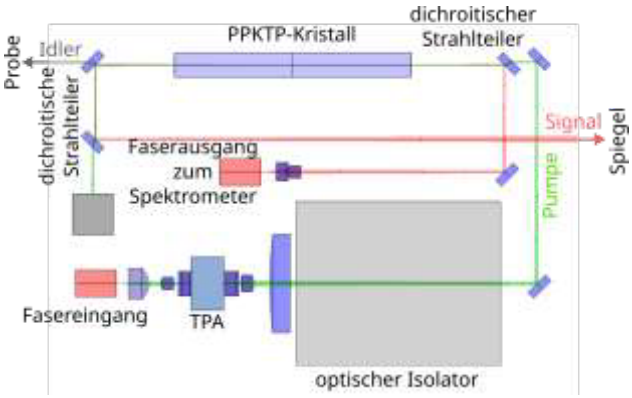


Bild 4 Schematische Darstellung der Strahlengänge auf der mikro-optischen Bank innerhalb des Mikromoduls.

Das Laser-System ist inklusive des TPAs integriert in eine Mikromodul, die notwendige Eingangsstrahlung des DBR-RWLs wird über einen Fasereingang zugeführt. Mittels Mikrolinsen wird die Strahlung aus der Faser dafür in den TPA gekoppelt. Nach der Verstärkung wird die Strahlung ebenfalls über Mikrolinsen kollimiert und durch eine sphärische Linse fokussiert. Nach Durchlaufen des optischen Isolators, der Umlenkspiegel und eines ersten dichroitischen Strahlteilers wird die Pumpe so in den PPKTP-Kristall eingekoppelt. Die erzeugte Idler-Strahlung passiert den zweiten dichroitischen Strahlteiler nach dem Kristall und steht als Freistrahl für die Interaktion mit der Probe zur Verfügung. Die ebenfalls generierte Signal-Strahlung wird hingegen von dem zweiten und dritten dichroitischen Strahlteiler reflektiert und verlässt die mikro-optische Bank in entgegengesetzter Richtung. Der für den interferometrischen Aufbau erforderliche Spiegel ist justierbar außerhalb der mikrooptischen Bank am Modulgehäuse angebracht.

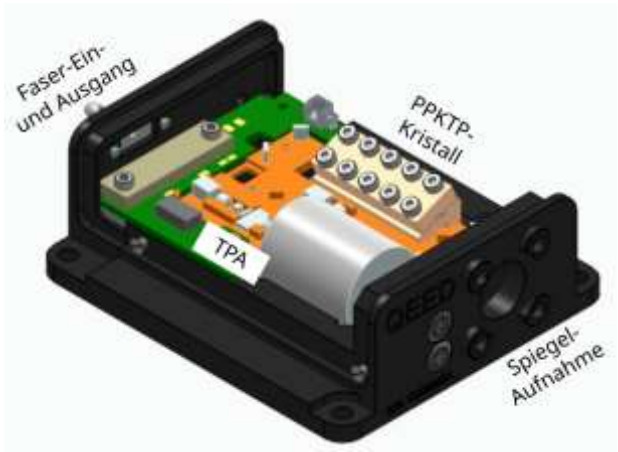


Bild 5 CAD-Zeichnung des mikrointegrierten Quantenlichtmoduls.

Die in Rückwärtsrichtung den Kristall verlassende Signal-Strahlung wird vom ersten dichroitischen Strahlteiler vor dem Kristall reflektiert und schließlich über Mikrolinsen in

DOI 10.5162/iCCC2026/P27

eine optische Faser eingekoppelt und kann so zu einem Spektrometer geführt werden.

Der gezeigte optische Aufbau ist mit Ausnahme des Kristalls auf einer mikro-optischen Bank untergebracht. Diese wird mittels Peltier-Elementen innerhalb des Modulgehäuses auf eine konstante Temperatur von 15°C stabilisiert. Eine CAD-Zeichnung des Gehäuses ist in **Bild 5** zu sehen, wobei das Gehäuse eine Grundfläche von lediglich 8,6 cm x 7,5 cm aufweist. Da der Kristall präzise in der Temperatur geregelt werden muss, ist dieser in einem separaten Kupfer-Halter mit eigener Temperaturregelung untergebracht. Die Spiegelaufnahme an der Außenseite des Modulgehäuses ermöglicht eine Justage des Spiegels für die Signal-Strahlung auch nach dem Zusammenbau. Dadurch kann das System auf unterschiedliche optische Weglängen zur Probe flexibel angepasst werden.

3.3 Messergebnisse

Die Sensing-Fähigkeiten des Quantenlichtmoduls lassen sich durch Detektion der Quanteninterferenz demonstrieren. Dazu wird die Idler-Strahlung durch einen Spiegel (also noch ohne Probe) und eine Linse in sich zurück reflektiert. Der Spiegel für die Signal-Strahlung wird in seiner optischen Weglänge entsprechend angepasst. Wenn sich die optischen Weglängen zwischen Idler- und Signal-Strahlung nur leicht unterscheiden, ist im Spektrum der Signal-Strahlung eine charakteristisches Interferenzmuster sichtbar. In **Bild 6** ist das gemessene Spektrum der Signal-Strahlung für diese Konfiguration gezeigt.

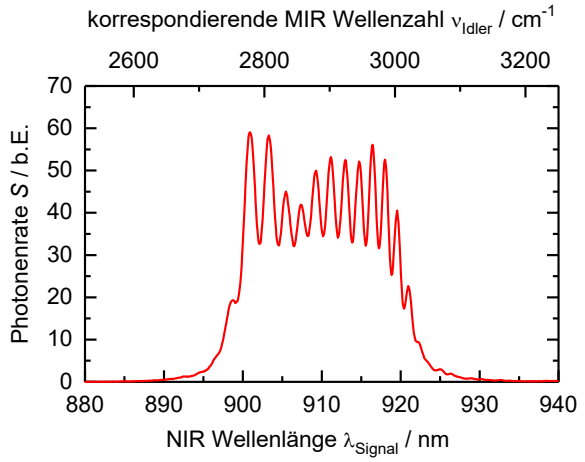


Bild 6 Gemessenes Spektrum der Signal-Strahlung des Quantenlichtmoduls mit sichtbarer Quanteninterferenz.

Das Spektrum der Signal-Strahlung erstreckt sich von etwa 900 nm bis 920 nm und zeigt die Quanteninterferenz in Form von starken Oszillationen. Aus dem Interferenzkontrast lässt sich die sogenannte Visibility auf etwa 30 % abschätzen, was mit den Werten aus Laborexperimenten gut übereinstimmt [3]. Da die Signal-Strahlung mit der Idler-Strahlung verschränkt ist, lässt sich aus dem Signal-Spektrum direkt das korrespondierende Spektrum der Idler-Strahlung angeben. Dies ist in der oberen horizontalen Achse in **Bild 6** gezeigt. Mit der mit dem Quantenlichtmodul erzeugten MIR-Strahlung lässt sich somit eine Probe im Wellenzahlenbereich von 2700 cm^{-1} bis 3100 cm^{-1} untersuchen.

4 Fazit

Das Konzept „Sensing mit nichtdetektiertem Licht“ konnte erfolgreich mit einem kompakten Quantenlichtmodul im MIR demonstriert werden. Dafür wurde eine Diodenlaserquelle bei 720 nm mit mehr 3 W optischer Ausgangsleistung entwickelt und das komplexe interferometrische Laborsetup weitestgehend miniaturisiert auf einer mikro-optischen Bank integriert. Das entwickelte Quantenlichtmodul hat eine Grundfläche von nur 8,6 cm x 7,5 cm und ermöglicht Spektroskopie in einem Wellenzahlenbereich von 2700 cm^{-1} bis 3100 cm^{-1}. Dies erlaubt insbesondere die spektroskopische Analyse der Valenzschwingungen von C-H_2-Gruppen innerhalb biochemischer Moleküle. Mit dem entwickelten Modul kommt die Spektroskopie im MIR mittels „Sensing mit nichtdetektiertem Licht“ Anwendungen einen entscheidenden Schritt näher. Gerade für die Krebsdiagnostik ergeben sich durch hyperspektrale Bildgebung enorme Potentiale.

5 Danksagung

Die Autoren danken dem Bundesministerium für Forschung, Technologie und Raumfahrt für die finanzielle Förderung im Rahmen des Verbundprojektes: „Quanten-Licht für eine neue Bildgebung zur frühen und differenzierten Erkennung von Biomarkern in der personalisierten Medizin (QEED)“ mit dem Förderkennzeichen 13N16381.

6 Literatur

[1] Kuepper, C et al.: Quantum Cascade Laser-Based Infrared Microscopy for Label-Free and Automated Cancer Classification in Tissue Sections. Sci Rep. 8, 2018, 10.1038/s41598-018-26098-w.

[2] Pilling, M et al.: Fundamental developments in infrared spectroscopic imaging for biomedical applications. Chem. Soc. Rev. 45, 2016, 10.1039/C5CS00846H.

[3] Kaufmann, P. et al: Mid-IR spectroscopy with NIR grating spectrometers, Opt. Express, 30, 2022, 10.1364/OE.442411.

[4] Vanselow, A. et al: Ultra-broadband SPDC for spectrally far separated photon pairs, Opt. Lett. 44, 2019, 10.1364/OL.44.004638.

[5] Mauerhoff, F. et al: GaAs Based Edge Emitters at 626 nm, 725 nm and 1180 nm, IEEE J. Select. Topics Quantum Electron. 31, 2025, 10.1109/JSTQE.2024.3431225.

DOI 10.5162/iCCC2026/P28

Explainability in ConvNeXt V2 for Contrast-Enhanced Spectral Mammography

Walid Ghorbel, Alireza Siyavashi, Christian Herglotz
Chair of Computer Engineering, Brandenburgische Technische Universität Cottbus–Senftenberg
Cottbus, Germany
E-Mail-Adresse: ghorbwal@b-tu.de, alireza.siyavashi@b-tu.de, christian.herglotz@b-tu.de

Abstract

Deep learning models can support breast cancer diagnosis in mammography, but their predictions are not explainable enough to count on their decision. Class activation mapping (CAM) techniques highlight image regions that contribute to a model's decision, yet existing approaches frequently produce diffuse saliency maps that include large areas with limited influence on the prediction. In this work, we employ Hybrid-CAM, a visualization method that combines global channel importance with high-resolution spatial activations, to generate more precise and faithful explanations of model behaviour. We show that even a comparatively simple convolutional neural network, when paired with an appropriate visualization technique, can yield substantially improved explainability.
Quantitative insertion–deletion metrics, together with qualitative visual inspection, demonstrate that Hybrid-CAM produces more focused and reliable explanations, particularly at intermediate network layers, by concentrating on regions that truly drive the model's predictions.

1 Introduction

Breast imaging plays a central role in detecting and assessing lesions, often guiding key diagnostic and treatment decisions [1]. In recent years, Artificial Intelligence (AI) has been increasingly used to support this process by analyzing medical images [2]. These models learn complex diagnostic patterns from large collections of medical images, enabled by their rich hierarchical representations [2].
Although they often achieve high performance in lesion detection and assessment, they usually provide only a final prediction without indicating which image regions contributed to it [1], [3]. This lack of transparency is especially problematic in breast imaging, where subtle or low-contrast abnormalities may influence clinical decisions [3].
Deep learning methods are also applied to Contrast-Enhanced Spectral Mammography (CESM), a radiographic modality designed to visualize cancerous lesions more clearly, to support lesion detection and diagnostic evaluation [4]. However, their predictions still do not indicate which image regions drive the outcome [3].
Clear visual explanations, complemented by quantitative assessment, are therefore essential for understanding model behavior and for analyzing potential errors [5].
To understand how these models make predictions, it is useful to consider how they process images. Most current methods for breast image analysis use convolutional neural networks (CNNs), which extract visual patterns through a series of layers. Each layer applies small learned filters that scan the image and produce feature maps that indicate where specific patterns occur [2] as shown in Figure 1.
In 2016, Zhou et al. introduced Class Activation Mapping (CAM), a technique that uses the final-layer feature maps to identify the image regions that contributed most to a given prediction [6]. CAM generates a spatial heatmap that highlights areas with high relevance for the model's decision, providing an intuitive visual representation of which regions influenced the outcome.

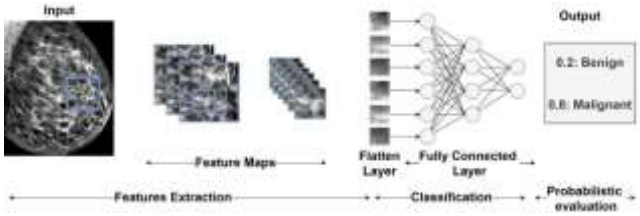


Figure 1: Overview of a typical CNN.

Several variants of CAM build on this idea. Gradient-weighted CAM (Grad-CAM) estimates how the model's output would change if a pattern detected in a deep feature map were strengthened, and uses this as an importance weight in the heatmap [7]. Because it operates on deep, low-resolution feature maps, the resulting heatmaps often highlight broad regions.
Score-CAM measures how the model's prediction changes when different parts of the input image are masked or revealed [8]. Regions that raise the class score when kept visible are considered important. Since each mask is tested independently, Score-CAM does not account for how multiple areas may jointly influence the prediction.
To address these limitations, we introduce Hybrid-CAM [9]. It starts from a Grad-CAM map and then evaluates which of the highlighted regions still support a high model prediction when the salient regions are selectively preserved in the image. Areas that continue to support the prediction are preserved, while diffuse or less relevant areas are suppressed. The resulting map isolates the areas that were most influential for the model's decision.
The main contributions of this work are as follows:

- We introduce Hybrid-CAM, a refinement of Grad-CAM that reduces over-highlighting and provides clearer, decision-relevant explanations.
- We provide a quantitative and qualitative comparison with Grad-CAM, Score-CAM, and Hybrid-CAM on a ConvNeXtV2-Tiny model [10]

DOI 10.5162/iCCC2026/P28

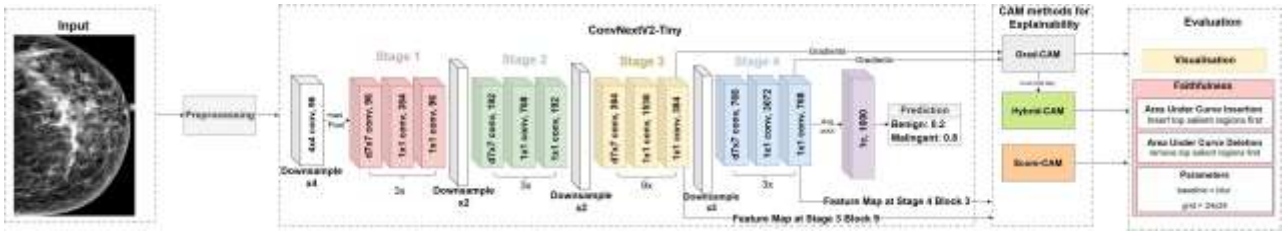


Figure 2: Overall workflow from input image and preprocessing, through ConvNeXtV2-Tiny feature extraction, to CAM generation (Grad-CAM, Hybrid-CAM, Score-CAM) and faithfulness evaluation.

using insertion-deletion metrics [11] on a public CDD-CESM dataset [12].

The remainder of this paper is organized as follows: Section 2 describes the dataset, backbone architecture, and the proposed Hybrid-CAM method. Section 3 outlines the evaluation protocol, and Sections 4 and 5 reports quantitative and qualitative results. In Section 6, we conclude the results and discuss future work.

2 Methodology

In this section, we outline how the fine-tuned ConvNeXtV2-Tiny model was prepared for explainability analysis on low-energy CESM images. After preprocessing, the images pass through the network, and feature maps are extracted from two layers: Stage 3 Block 9, the deepest layer that still preserves a clear spatial layout of the breast, and Stage 4 Block 3, the final block before classification. These maps are used to compute Grad-CAM, Score-CAM, and Hybrid-CAM. The resulting explanations are then evaluated with insertion–deletion metrics and visualized on representative test cases.

2.1 Dataset and Preprocessing

Experiments were conducted on the low-energy subset of the public CDD-CESM dataset [12]. A ConvNeXtV2-Tiny model pretrained on ImageNet-22K was fine-tuned for benign–malignant classification [10], achieving 92.2% specificity, 56.1% sensitivity, and 82.1% precision. CAM analyses were performed on a held-out test set of 141 images, following the pipeline illustrated in Fig. 2. All test images were processed using the same validation pipeline used during model training.

2.2 Backbone Network: ConvNeXtV2-Tiny

ConvNeXtV2-Tiny is used as the backbone in this study. It is a convolutional network that integrates several design elements inspired by transformer models, while retaining the spatially organized processing characteristic of CNNs [10]. This combination allows the model to capture both local details and broader patterns, which are crucial in breast imaging, where the appearance of a lesion often depends on its surrounding tissue. Figure 2 illustrates the four stages of the ConvNeXtV2 backbone, where earlier stages emphasize low-level image structures such as edges or subtle textures, while deeper stages capture complex diagnostic patterns. Each stage contains several ConvNeXt blocks, which act as small processing units that extract and refine patterns using a 7×7 depthwise convolution to examine a larger local area, followed by two 1×1 convolutions that combine and refine the extracted features. Across stages, the number of feature maps increases, allowing the model to capture more complex patterns. For the explainability analysis, we use the last block of Stage 3 and the last block of Stage 4 of the ConvNeXtV2-Tiny model, as shown in Fig. 2. At these depths, the feature maps retain sufficient spatial resolution to indicate where the model is focusing, while also encoding higher-level information that strongly influences the benign-malignant decision. These characteristics make these blocks well-suited for CAM-based explanation and for assessing whether the model relies on clinically relevant regions.

2.3 Grad-CAM

CAM methods produce spatial heatmaps that highlight which regions the model considers relevant for a prediction [6]. Grad-CAM does this by using the gradients of the class score with respect to the feature maps of a selected layer [7]. Let $\boldsymbol{A_k}$ denote the $\boldsymbol{k}$-th feature map at this layer, with spatial dimensions $\boldsymbol{H \times W}$. A feature map is the 2D activation map produced by one channel of a convolutional layer, where each value $\boldsymbol{A_k(i,j)}$ indicates how strongly this channel responds to a pattern at the location $\boldsymbol{(i,j)}$. Let $\boldsymbol{y_c}$ denote the model score for class $\boldsymbol{c}$ before the final softmax. The gradient $\boldsymbol{\partial y_c / \partial A_k(i,j)}$ describes how much the class score would change if the activation $\boldsymbol{A_k(i,j)}$ were increased, and thus reflect its relevance for class c. Grad-CAM then computes a single importance weight for each feature map by averaging these gradients over all spatial positions: $\boldsymbol{w_k} = \sum_{i,j} \frac{\partial y_c}{\partial A_{k(i,j)}}$

A positive w_k means that increasing feature map k tends to increase the prediction for class c, whereas a negative w_k indicates the opposite.

The Grad-CAM heatmap is then obtained as:

$$CAM(i,j) = ReLU\left(\sum_k w_k A_k(i,j)\right)$$

where the summation aggregates contributions from all feature maps. The ReLU (rectified linear unit) function sets negative values to zero, so that only features that support the prediction remain visible in the map. The heatmap is finally upsampled and normalized for visualization.

2.4 Hybrid-CAM

Grad-CAM shows where the model focuses, but it does not reveal how strongly these highlighted regions influence the

prediction when those regions are actually perturbed or removed. Hybrid-CAM addresses this by introducing controlled perturbations: it preserves only the most salient areas of the Grad-CAM map, replaces the remaining pixels with a baseline, and measures how the model's output changes. Given an input image x and its Grad-CAM map $A(i,j)$, pixels are ranked by saliency. For each area level $p \in \{1,2,3,5,10,20,35,50,70,90\}\%$, indicating that only the top p of $A(i,j)$ is retained, a binary mask M_k preserves those locations while the remaining pixels are replaced with a blurred baseline x_{blur} (Gaussian blur, σ =35):

$$x_p = M_p \odot x + (1 - M_p) \odot x_{blur}$$

where $\odot$ denotes elementwise multiplication across spatial locations.

Let $y(\cdot)$ denote the model's output score for class c, i.e., the raw value produced by the final linear layer before probabilities are computed. This score can take any real value, with higher values indicating stronger evidence for the class c. For each masked input x_p we compute

$$\Delta(p) = max\left(0,\, y(x_p) - y(x_{blur})\right)$$

which measures how strongly the preserved region increases the prediction compared to the blurred baseline.

These effects are projected back to the spatial domain via

$$S(i,j) = \sum_p w(p)\ \Delta(p)\, M_p(i,j), \quad w(p) = \frac{1}{\sqrt{p}}$$

Where $S(i,j)$ reflects the accumulated contribution of the preserved regions across different mask sizes. The factor $w(p) = 1/\sqrt{p}$ slightly down-weights masks with larger area levels p, which would otherwise add more to $S(i,j)$ simply because they cover more pixels. This keeps the influence of small, focused masks and larger masks on a comparable scale.

The resulting map is combined with the Grad-CAM map to obtain the final Hybrid-CAM heatmap:

$$H(i,j) = norm\big(A(i,j) \odot S(i,j)\big)$$

H highlights locations that are strongly emphasized by Grad-CAM and that still increase the class score when they are preserved in the perturbed inputs x_p. This focuses the heatmap on regions that make a robust positive contribution to the prediction, rather than on broad areas with weak or diffuse responses.

3 Faithfulness Evaluation

We assess explanation quality using the insertion–deletion (I/D) framework from RISE (Randomized Input Sampling for Explanation) [11], a widely used faithfulness metric for saliency maps [5]. Each image is divided into a $g \times g$ grid, meaning it is split into g rows and g columns of equally sized patches. For every patch, we compute a relevance score as the average value of the explanation map $A(i,j)$ inside that patch. Sorting these patch scores then gives an ordering from most to least relevant. Masked patches are replaced with a Gaussian-blurred baseline image σ =35. For a fraction p of inserted or removed patches, we measure the model's score, denoted $P_{\text{ins}}(p)$ and $P_{\text{del}}(p)$. Faithfulness is summarized by the area under these curves:

$$\text{AUC}_{\text{ins/del}} = \int_0^1 P_{\text{ins/del}}(p)\ dp$$

A high AUC_{ins} indicates that relevant regions are added early, whereas a low AUC_{del} indicates that removing highlighted regions strongly reduces the prediction. We vary the explanation layer (Stage 3 Block 9 and Stage 4 Block 3). The perturbation grid was fixed to $g = 24 \times 24$, and all metrics used the model's predicted class as the target to evaluate faithfulness to its own decision.

4 Quantitative Results

	Score-CAM $\text{AUC}_{\text{ins}}/\text{AUC}_{\text{del}}$	Grad-CAM $\text{AUC}_{\text{ins}}/\text{AUC}_{\text{del}}$	Hybrid-CAM $\text{AUC}_{\text{ins}}/\text{AUC}_{\text{del}}$
TP	0.407/0.955	0.958/0.361	0.963/0.350
TN	0.941/0.830	0.931/0.727	0.942/0.742

Table 1: Average Insertion and Deletion AUC for Score-CAM, Grad-CAM, and Hybrid-CAM at Stage 3 Block 9, computed with a blur baseline (σ = 35) and a 24×24 grid. Values are averaged over TP (32 cases) and TN (83 cases).

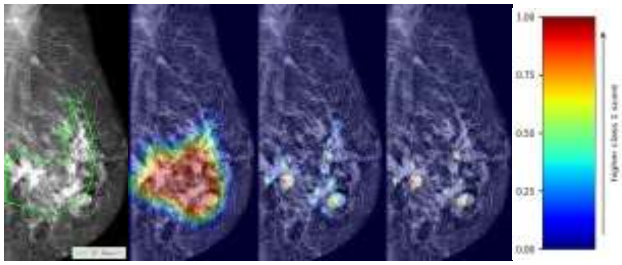
	Score-CAM $\text{AUC}_{\text{ins}}/\text{AUC}_{\text{del}}$	Grad-CAM $\text{AUC}_{\text{ins}}/\text{AUC}_{\text{del}}$	Hybrid-CAM $\text{AUC}_{\text{ins}}/\text{AUC}_{\text{del}}$
TP	0.911/0.254	0.593/0.737	0.612/0.735
TN	0.957/0.879	0.906/0.888	0.827/0.928

Table 2: Average Insertion and Deletion AUC at Stage 4 Block 3.

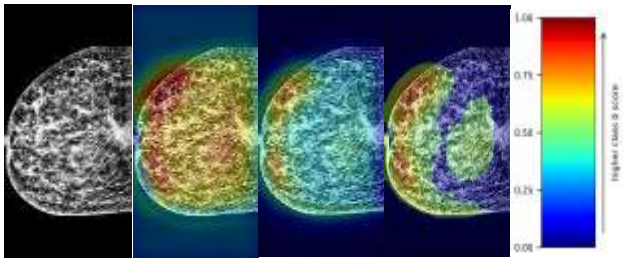
For true positives (TP) in Stage 3 Block 9 as shown in Table 1, Hybrid-CAM achieves the strongest results (0.963/0.350), closely followed by Grad-CAM (0.958/0.361). Both methods clearly affect the model's prediction when their highlighted regions are modified, suggesting that they capture the area that are most relevant for the decision. Score-CAM performs weakly (0.407/0.955), meaning that modifying its highlighted regions has little impact on the output. Table 2 shows the results in Stage 4 Block 3, the scores for Hybrid-CAM and Grad-CAM become weaker, with lower insertion and higher deletion values. This indicates that changes in the highlighted regions have a less direct effect on the prediction at this depth. For true negatives, all methods show higher deletion values, which is expected because benign predictions do not rely on a single dominant region but rather on the overall absence of suspicious patterns. Overall, Hybrid-CAM provides slightly more faithful explanations than Grad-CAM at Stage-3 Block-9. In contrast, Score-CAM performs best at the deepest layer (Stage-4 Block-3).

5 Qualitative Results

To complement the quantitative findings, we qualitatively inspected the explanations for a representative true-positive and true-negative example Fig. 3a and Fig. 3b. Hybrid-CAM and Grad-CAM are shown at Stage 3 Block 9, where they exhibited the strongest quantitative effect. Score-CAM is shown at Stage 4 Block 3. In these visualizations, the color scale indicates the relative contribution to the predicted class score: for malignant cases, the maps show contributions to the class-1 score, while for benign cases, they show contributions to the class-0 score.



Original, Score-CAM, Grad-CAM, Hybrid-CAM
(a) True positive



Original, Score-CAM, Grad-CAM, Hybrid-CAM
(b) True negative

Figure 3: Qualitative comparison of Score-CAM, Grad-CAM, and Hybrid-CAM for a TP and TN example. Each block shows the original image and the corresponding saliency maps, with a separate color bar indicating the score scale for the respective class.

Fig. 3a shows a true-positive mammogram case. Both Hybrid-CAM and Grad-CAM highlight a small number of distinct regions that contribute to the malignant prediction. Grad-CAM displays several separate activation points, whereas Hybrid-CAM produces a more compact pattern in similar locations. Score-CAM shows a broader activation pattern in this deeper layer, which matches its stronger quantitative influence at Stage 4 Block 3.

Fig. 3b illustrates a true-negative example. The highlighted regions mainly contribute to the benign prediction. Hybrid-CAM and Grad-CAM show low-intensity responses, often around breast edges and broader texture patterns, suggesting that the model does not rely on a single dominant region for this decision. Score-CAM again produces a wider activation pattern at Stage 4 Block 3, consistent with the more diffuse behavior observed in the quantitative evaluation for this layer.

Overall, the qualitative examples illustrate different activation behaviors within the selected layers, with Hybrid-CAM and Grad-CAM showing more compact patterns at Stage 3 Block 9 and Score-CAM appears more diffuse in Stage 4 Block 3 used for visualization.

6 Conclusion

This work investigated explainability in a CNN-based classifier using Class Activation Mapping techniques to explain the model decision using saliency maps. We introduced Hybrid-CAM, a refinement of Grad-CAM that incorporates controlled perturbations to assess whether highlighted regions truly affect the model output. Quantitative insertion-deletion metrics and qualitative examples showed that Hybrid-CAM produces more focused and faithful explanations than Grad-CAM and Score-CAM, particularly at intermediate network layers. These results indicate that evaluating the influence of highlighted regions can provide clearer insight into how deep learning models form their predictions in breast imaging.

In future work, we will take the main feature of ConvNextv2 and modify each of these features to find out how harsh the result will change. It will clarify for the novice medical practitioner to learn to identify cancerous lesions.

7 References

[1] L. Shen, L. R. Margolies, J. H. Rothstein, E. Fluder, R. McBride, and W. Sieh, “Deep Learning to Improve Breast Cancer Detection on Screening Mammography,” *Sci Rep*, vol. 9, no. 1, p. 12495, Aug. 2019, doi: 10.1038/s41598-019-48995-4.

[2] L. Alzubaidi *et al.*, “Review of deep learning: concepts, CNN architectures, challenges, applications, future directions,” *J Big Data*, vol. 8, no. 1, p. 53, 2021, doi: 10.1186/s40537-021-00444-8.

[3] D. K. Gurmessa and W. Jimma, “Explainable machine learning for breast cancer diagnosis from mammography and ultrasound images: a systematic review,” *BMJ Health Care Inform*, vol. 31, no. 1, p. e100954, Feb. 2024, doi: 10.1136/bmjhci-2023-100954.

[4] C. Y. Y. Wong, S. Y. S. Lee, and R. D. Mahmood, “Contrast-enhanced spectral mammography,” *Singapore Med J*, vol. 65, no. 3, pp. 195–201, Mar. 2024, doi: 10.4103/singaporemedj.SMJ-2021-268.

[5] T. Gomez, T. Fréour, and H. Mouchère, “Metrics for saliency map evaluation of deep learning explanation methods,” 2022, *arXiv*. doi: 10.48550/ARXIV.2201.13291.

[6] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba, “Learning Deep Features for Discriminative Localization,” 2015, *arXiv*. doi: 10.48550/ARXIV.1512.04150.

[7] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, “Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization,” 2016, doi: 10.48550/ARXIV.1610.02391.

[8] H. Wang *et al.*, “Score-CAM: Score-Weighted Visual Explanations for Convolutional Neural Networks,” 2019, *arXiv*. doi: 10.48550/ARXIV.1910.01279.

[9] Christian Herglotz, Alireza Siyavashi, “Toward an Energy-Efficient and Explainable Neural Network Architecture for Detection of Breast Cancer in Mammography,” *Brandenburgische Technische Universität Cottbus–Senftenberg*, 2023.

[10] S. Woo *et al.*, “ConvNeXt V2: Co-designing and Scaling ConvNets with Masked Autoencoders,” 2023, *arXiv*. doi: 10.48550/ARXIV.2301.00808.

[11] V. Petsiuk, A. Das, and K. Saenko, “RISE: Randomized Input Sampling for Explanation of Black-box Models,” 2018, *arXiv*. doi: 10.48550/ARXIV.1806.07421.

[12] R. Khaled *et al.*, “Categorized Digital Database for Low energy and Subtracted Contrast Enhanced Spectral Mammography images.” The Cancer Imaging Archive, 2021. doi: 10.7937/29KW-AE92.

DOI 10.5162/iCCC2026/P29

Effekte der Quantisierung auf Retrieval-Augmented-Generation Language Modelle für die Parkinson-Versorgung

Patrick Gaudl, Christoph-Alexander Holst und Volker Lohweg, Institut für industrielle Informationstechnik (inIT), Technische Hochschule Ostwestfalen-Lippe, Lemgo, Deutschland, vorname.nachname@th-owl.de

Kurzfassung

Large Language Models bieten großes Potenzial für assistive Anwendungen in der Parkinsonversorgung, sind jedoch aufgrund ihres hohen Speicherbedarfs bislang schwer in der Praxis einsetzbar. Dieser Beitrag untersucht den Einfluss verschiedener Quantisierungsmethoden auf ein Retrieval-Augmented-Generation-System auf Basis eines Qwen-Modells. Die Ergebnisse zeigen, dass sich der Speicherbedarf auf bis zu 36 % reduzieren lässt, bei leichter bis moderater Beeinträchtigung der Antwortqualität. Die Ergebnisse demonstrieren, dass lokal ausführbare Parkinson-Assistenzsysteme technisch realisierbar sind.

Abstract

Large Language Models hold significant potential for assistive applications in Parkinson's care, yet their substantial memory requirements currently limit deployment in real-world clinical environments. This paper investigates the impact of different quantization methods on a Retrieval-Augmented Generation system based on a Qwen model. Our results show that model size can be reduced to as little as 36% of the original footprint, with only mild to moderate degradation in response quality. These findings demonstrate that locally executable, Parkinson's-specific assistant systems are technically feasible and enable privacy-preserving assistance systems.

1 Motivation

Parkinson ist die weltweit am schnellsten zunehmende neurodegenerative Erkrankung und geht mit einer Vielzahl motorischer und nichtmotorischer Symptome einher [1]. Durch die Heterogenität der Symptomprogression sowie die oft fragmentierte Versorgung entstehen erhebliche Herausforderungen bei der Dokumentation, Bewertung und Therapieanpassung [2].

Digitale Assistenzsysteme bieten hier einen Ansatz zur strukturierten Aufbereitung individueller Krankheitsverläufe und können damit Fachkräfte im Gesundheitswesen sowie Betroffene unterstützen [3].Gleichzeitig stehen gegenwärtige Large Language Models (LLMs) aufgrund ihres enormen Speicher- und Rechenbedarfs einer breiten Anwendung in der klinischen Praxis entgegen. Die Ausführung solcher Modelle erfolgt typischerweise auf leistungsfähiger Cloud-Infrastruktur. Für den Einsatz in hochsensiblen, personenbezogenen Anwendungsszenarien wie der Parkinson-Versorgung ist eine lokale Verarbeitung auf Endgeräten im Hinblick auf Datenschutz, Verfügbarkeit und regulatorischer Anforderungen besonders attraktiv.

Damit LLM-basierte Assistenzsysteme im Routinebetrieb der Parkinsonversorgung ankommen können, müssen (1) der Speicherbedarf großer Modelle drastisch reduziert werden und dabei (2) gleichzeitig die Antwortqualität gewahrt werden. Quantisierungsmethoden ermöglichen eine erhebliche Verringerung des Modell-Memory-Footprints, ohne das Modell neu trainieren zu müssen [4]. Für domänenspezifische Assistenzsysteme ist zusätzlich eine Wissensintegration erforderlich, um Halluzinationen zu minimieren und medizinische Aussagequalität sicherzustellen. *Retrieval-Augmented-Generation* (RAG) hat sich hierfür als vielversprechende Methode etabliert [5].

Vor diesem Hintergrund untersucht dieser Beitrag den Einfluss unterschiedlicher Quantisierungsverfahren auf ein RAG-basiertes LLM im Kontext der Parkinsonversorgung. Dazu wird ein offenes Modell der Qwen-Familie in mehreren Quantisierungsstufen evaluiert und auf eigens kuratierten Frage-Antwort-Datensätzen aus Leitlinieninhalten getestet. Ziel ist die Beantwortung der Kernfrage, wie stark ein LLM quantisiert werden kann, bevor relevante Einbußen in der Ausgabequalität auftreten.

2 Stand der Forschung

2.1 LLMs in Medizin- und Gesundheitswesen

Generative künstliche Intelligenz und LLMs eröffnen zunehmend neue Anwendungsmöglichkeiten im Gesundheitswesen, insbesondere im klinischen Informationsmanagement [6]. LLMs können an verschiedenen Stellen der Versorgung eingesetzt werden, etwa zur Unterstützung klinischer Entscheidungsprozesse [7], zur Automatisierung medizinischer Dokumentation [8], zur Analyse von Fachliteratur oder als Bestandteil patientenorientierter Assistenzsysteme [3].

Gleichzeitig bestehen wesentliche Herausforderungen für den Einsatz im klinischen Kontext. Allgemeine LLMs verfügen über keine explizite domänenspezifische Wissensbasis und interpretieren medizinische Fachterminologie mitunter unvollständig, was ihre Leistungsfähigkeit in spezialisierten Aufgaben einschränkt [6]. Zudem sind LLMs anfällig für Halluzinationen, Jailbreak-Prompts und Datenvergiftungsangriffe. Forschungsarbeiten zu Unsicherheitsabschätzung [3], Modell-Governance und hybride Wissenszugriffstechniken adressieren diese Risiken, befinden sich jedoch noch in einem frühen Entwicklungsstadium

DOI 10.5162/iCCC2026/P29

[9]. Damit LLMs im medizinischen Alltag zuverlässig eingesetzt werden können, sind Mechanismen erforderlich, die domänenspezifisches Wissen integrieren, Fehlverhalten reduzieren und nachprüfbare Aussagen ermöglichen.

2.2 Retrieval-Augmented Generation und Fine-Tuning

RAG und Fine-Tuning gehören zu den zentralen Methoden, um allgemeine Sprachmodelle für medizinische Spezialaufgaben anzupassen. Die Ansätze ermöglichen es, LLMs mit einer Wissensbasis zu verknüpfen und so natürlichsprachliche Interaktion über medizinische Dokumente hinweg zu unterstützen [5], [10]. Der RAG-Prozess gliedert sich in: (1) Indexing, bei dem Textquellen in kleinere Einheiten zerlegt und mittels Embedding-Modellen als Vektoren in einem Wissensspeicher abgelegt werden, (2) Retrieval, bei dem Nutzereingaben als Embeddings kodiert und die semantisch ähnlichsten Wissenseinheiten identifiziert werden, sowie (3) Generation, in welcher das LLM die abgerufenen Inhalte in seine Antwort integriert [5].
RAG reduziert Halluzinationen, indem es die Ausgabe auf verifizierbare Quellen stützt. Fine-Tuning nutzt zusätzliche domänenspezifische Trainingsdaten, wodurch Modelle medizinische Terminologie präziser interpretieren können. Erste prototypische Systeme wie Alpacare [10] oder erste Parkinson-spezifische LLM-Implementationen [11] decken bislang nur eng definierte Fragestellungen ab. Offene Forschungsfragen bestehen hinsichtlich der Skalierbarkeit, der Robustheit gegen Angriffe und den Grenzen komprimierter Modellausführungen im klinischen Umfeld.

2.3 Quantisierung von LLMs

LLMs sind während ihres Trainings und der Inferenz oft mit hohen Anforderungen an Hardware bezüglich Speicher und Rechenleistung verbunden, da sie eine hohe Anzahl trainierbarer Parameter beinhalten [4]. Um dieses Problem zu lindern und die Hardwareanforderungen von LLMs zu reduzieren, dient die Modellquantisierung. Dabei handelt es sich um die Skalierung der jeweiligen Modellparameter von ihrem ursprünglichen Datentyp auf einen Datentyp mit reduziertem Speicherbedarf [4].
Die Quantisierung eines Tensors lässt sich am Beispiel der *absmax*-Quantisierung [12] darstellen: Seien $\boldsymbol{X}_{FP16}$ der als *16-Bit Floating Point (FP16)* vorliegende unquantisierte Tensor und $s_{x,FP16}$ die für ein Skalierungsfaktor, so lässt sich die Repräsentation ebendieses Tensors $\boldsymbol{X}_{Int8}$ mit *8-Bit-Integer (Int8),* wie folgt, berechnen [12]:

$$\boldsymbol{X}_{\text{Int8}} = round\left(\frac{127}{max_{ij}(|\boldsymbol{X}_{\text{FP16}}|)} \cdot \boldsymbol{X}_{\text{FP16}}\right) = round\left(s_{x,\text{FP16}} \cdot \boldsymbol{X}_{\text{FP16}}\right).$$

Dadurch ergibt sich eine Integer-Repräsentation des Tensors, dessen Elemente sich im von Int8 abgebildeten Wertebereich von $[-127{,}127]$ befinden. Der dequantisierte Tensor $\boldsymbol{X}_Q$ wird anschließend wie folgt berechnet [12]: $\boldsymbol{X}_Q = S_{\text{FP16}} \cdot \boldsymbol{X}_{\text{Int8}} = \frac{1}{s_{x,FP16}} \cdot \boldsymbol{X}_{\text{Int8}}$. Bei der Dequantisierung können durch Rundung Quantisierungsfehler E_Q auftreten [13]: $E_Q = \boldsymbol{X}_{\text{FP16}} - \boldsymbol{X}_Q$. Quantisierungsfehler können sich auf folgende Berechnungen auswirken und somit die Leistung des Modells beeinflussen [13], weshalb diese minimiert werden sollten.

3 Methodik

3.1 Systemarchitektur

3.1.1 RAG-Pipeline

Die Implementierung (Bild 1) umfasst eine RAG-Architektur, welche mithilfe des Frameworks *Langchain* realisiert wird. Für die Implementierung der Wissensbasis als Vektordatenbank wird *Facebook AI Similarity Search (FAISS)* genutzt. Für das Erstellen der Embeddings wurde der Sentence-Transformer *all-MiniLM-L6-v2* verwendet.

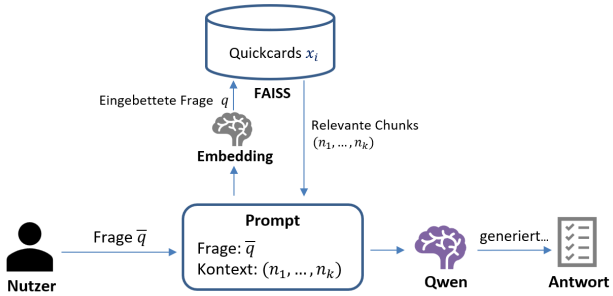


Bild 1 Aufbau des RAG-Systems: Die Frage $\overline{q}$ wird dem Prompt zugeführt und per Embedding in ihre Vektorrepräsentation umgewandelt, welche als Anfrage an die Vektordatenbank dient. Die k abgerufenen Dokumente werden dem Prompt als Kontext zugeführt.

Als Sprachmodell dient das von der Alibaba-Group entwickelte *Qwen2.5-7B-Instruct*. Das Modell umfasst 7,61 Milliarden Parameter und ist auf dem Open-LLM-Leaderboard (huggingface.co; Stand Februar 2025) als eines der beliebtesten Modelle vertreten.
Bei der Implementierung wird eine Temperatur von $t = 10^{-2}$ verwendet, um die Zufälligkeit der generierten Antworten gering zu halten. Zudem wurde dem Modell durch einen System-Prompt die Anweisung gegeben, möglichst kurz anhand des bereitgestellten Kontexts zu antworten, um den Spielraum für Halluzinationen zu minimieren.

3.1.2 Leitlinien als Wissensbasis

Bei den für die Vektordatenbank verwendet Inhalten handelt es sich um sogenannte *Quickcards*, die vom *Parkinsonnetz Münsterland+* herausgegeben wurden, auf aktuellen Parkinson-Leitlinien basieren und Bereiche, wie die Logopädie, Physiotherapie, Ergotherapie und die Neurologie abbilden. Konkret handelt es sich dabei um Symptome und die dazugehörigen Empfehlungen. Die initial als PDF vorliegenden Quickcards wurden per Hand in Textform überschrieben und im *MarkDown*-Format eingespeichert.
Bei der Einspeisung in die Wissensbasis wird auf eine fixe Chunk-Größe verzichtet, da sich die Länge der Einträge zu verschiedenen Symptomen deutlich voneinander unterscheiden und der Abbruch von Sätzen sowie das Einbringen kontextfremder Informationen verhindert werden soll.

3.1.3 Quantisierungsstufen

Das Modell wurde in drei verschiedenen Quantisierungsstufen untersucht. Die Modellparameter liegen initial im Datenformat *BrainFloat16 (BF16)* [14] vor, wobei jeder Modellparameter 16 Bits Speicherplatz benötigt. Als Quantisierungen werden die von Dettmers et al. vorgestellten Algorithmen *LLM.Int8* [12] und *4-bit block-wise quantization* [15] verwendet..

Quantisierungsstufe	Datentyp	Speicherbedarf*
Keine	BF16	15.23 GB
LLM.Int8	Int8	8.7 GB
4-bit block-wise	NF4	5.44 GB

Tabelle 1 Übersicht über die verwendeten Quantisierungsmethoden und ihre Auswirkungen auf den Speicherbedarf (VRAM) des Modells.

3.2 Evaluation

Die Auswirkungen der jeweiligen Quantisierungsstufen wurden anhand von drei Evaluationsdatensätze getestet und anhand verschiedener Metriken evaluiert. Die bei der Evaluation verwendeten *Quickcard-QA (Question-Answering)*-Datensätze beziehen sich inhaltlich auf die als Wissensbasis verwendeten Quickcards und orientieren sich in ihrem Aufbau an gängigen Benchmarks wie z.B. *PubMedQA* [16]. So handelt es sich bei *MC (Multiple-Choice)-QQA-1* und *-2)* um Multiple-Choice-Fragen mit direkten symptombezogenen Fragen, bei denen das Modell zwischen Antwortmöglichkeiten auswählen muss, und bei *Gen (Generative)-QQA* um offene Fragen, bei denen das Modell Antwortsequenzen generieren muss.

Datensatz	Fragentyp	Metriken
MC-QQA-1	Single/Multiple-Choice QA, direkte symptombezogene Fragen	Accuracy, Precision, Recall, F1
MC-QQA-2	Single/Multiple-Choice QA mit fiktiven Fallbeispielen	Accuracy, Precision, Recall, F1
Gen-QQA	Generative QA mit fiktiven Fallbeispielen	Accuracy durch menschliche Evaluation, BERTScore

Tabelle 2 Übersicht über die drei verwendeten Evaluationsdatensätze, deren verwendeten Fragentypen und deren angewendete Evaluationsmetriken.

Die Antworten der Modelle auf die Fragen aus MC-QQA-1 und -2 wurden anhand der Klassifikationsmetriken *Accuracy*, *Precision*, *Recall* und *F1-Score* evaluiert. Für die Evaluation der Frei-Text-Modellantworten für Gen-QQA wurde sowohl eine menschlichen Beurteilung als auch der *BERTScore [17]* genutzt. BERTScore vergleicht die Token des zu prüfenden Satzes mit denen des Referenzsatzes und berechnet darauf basierend die Metriken Precision (P_{BERT}), Recall (R_{BERT}) und einen F1-Score (F_{BERT}) [17].

Für die Evaluation wurde eine *Nvidia Titan RTX* mit 24 GB Grafikspeicher und einer Leistung von 130 Tensor TFLOPS und ein *AMD Ryzen 7 3800X 8-Core*-Prozessor verwendet. Für jede neue Frage wurde ein neuer Kontext eröffnet, um keine Informationen aus den vorherigen Fragen einfließen zu lassen.

4 Ergebnisse

Tabelle 3 zeigt auf, dass Int8 über alle Datensätze hinweg die höchsten Accuracy-Scores und den höchsten F_{BERT} liefert. Die Quantisierung mit NF4 zeigt eine leichte Performance-Degradation, während die Leistung von BF16 nahe an der von Int8 liegt.

Datensatz	Metrik	BF16	Int8	NF4
	Speicherbedarf	15.23 GB	8.7 GB	5.4 GB
MC-QQA-1	Accuracy	0.757	**0.779**	0.764
	Precision	0.657	**0.691**	0.662
	Recall	0.891	**0.905**	0.843
	F1	0.733	**0.76**	0.724
MC-QQA-2	Accuracy	0.71	**0.75**	0.68
	Precision	0.55	**0.6**	0.547
	Recall	**0.78**	**0.78**	0.76
	F1	0.621	**0.659**	0.608
Gen-QQA	Accuracy*	**0.6**	0.52	0.4
	P_{BERT}	0.673	**0.678**	0.632
	R_{BERT}	0.759	**0.762**	0.717
	F_{BERT}	0.712	**0.717**	0.671

Tabelle 3 Übersicht der über den jeweiligen Datensatz gemittelten Evaluationsergebnisse des Qwen2.5-7B-Instruct in den verwendeten Quantisierungsstufen. *Bewertung durch menschlichen Experten

Die Recall-Scores im Falle von MC-QQA-1 und -2 liegen zudem deutlich über den Precision-Scores. Dies ist ein Indiz dafür, dass die Modelle viele tatsächlich falsche Antworten als richtig interpretieren, was im Rahmen einer medizinischen Assistenzanwendung problematisch wäre.
Ferner fällt auf, dass sich bei Gen-QQA die BERTScore-Metriken nur leicht voneinander unterscheiden, während die Evaluation durch Menschen eine deutliche Leistungsdegradation mit steigender Quantisierungsstufe feststellt.
Allgemein lassen sich zwischen den einzelnen Quantisierungsstufen nur geringe Abweichungen beobachten, wobei diese bei NF4 etwas stärker ausgeprägt sind. Das beste Verhältnis zwischen Speicherplatz und Leistung liefert jedoch das Modell mit Int8-Quantisierung.

5 Diskussion

5.1 Limitationen

Die Bewertung der generierten Antworten erfolgte notwendigerweise über qualitative Beurteilungen, bei denen individuelle Einschätzungen eine Rolle spielen. Solche Einschätzungen sind in domänenspezifischen Benchmarks üblich, beinhalten jedoch unvermeidbare subjektive Komponenten, die in zukünftigen Arbeiten durch Mehrgutachter oder ergänzende automatische Metriken weiter objektiviert werden können.
Als Wissensbasis dienten Quickcards, die eine kontrollierte, aber vereinfachte Repräsentation klinischer Informationen darstellen. Komplexe Versorgungssituationen,

wie Mehrfacherkrankungen, Kontraindikationen oder vollständige Patientenhistorien, wurden nicht abgebildet.
Schließlich bestehen systembedingte Limitierungen in der technischen Pipeline. Das verwendete Embedding-Modell besitzt eine begrenzte Kontextlänge und ist nicht auf deutsche medizinische Terminologie spezialisiert, was das Retrieval in Einzelfällen beeinträchtigen kann. Auch das generative Modell zeigt gelegentliche Unsicherheiten im Umgang mit fachspezifischen Begriffen, was weiteres Fine-Tuning nahelegt.

5.2 Fazit

Die Arbeit zeigt, dass sich die Quantisierung als bewährtes Mittel zur ressourcenschonenden Implementierung von RAG-Systemen bewährt, da die Leistungseinbußen mit zunehmender Stärke der Quantisierung moderat ausfallen. Selbst die stärkste Quantisierungsstufe, die den Speicherbedarf des Modells um fast Dreiviertel senkt, geht mit nur einem moderaten Leistungsabfall einher. Ebenfalls unterstreichen die Ergebnisse, dass sich RAG als geeignete Methode bewährt, um LLMs bei der Generation kuratiertes Fachwissen zur Verfügung zu stellen. Durch die Kombination von geeigneten Quantisierungsmethoden und einem RAG-System mit einer kuratierten Wissensbasis, lassen sich fachlich hochspezialisierte LLMs auch auf handelsüblichen Grafikkarten ausführen.

5.3 Ausblick

Zur Verbesserung des Retrievals bieten sich weiterentwickelte RAG-Verfahren wie *Query-Rewriting*, modulare RAG-Architekturen und LLM-basierte Bewertungsinstanzen an. Ein domänenspezifischer Ausbau des Systems stellt weiteres Potenzial dar. Fine-Tuning mit medizinischen Datensätzen – beispielsweise mittels QLoRA – sowie der Einsatz spezialisierter deutscher Embedding-Modelle können die Robustheit der Generierung und Terminologieverarbeitung erhöhen. Darüber hinaus eröffnen Multi-Agenten-Ansätze und Reasoning-Strategien wie Chain-of-Thought neue Möglichkeiten, komplexere klinische Fragestellungen abzubilden und die Qualität datenbasierter Entscheidungsunterstützung weiter zu steigern.

6 Literatur

[1] D. Su *et al.,* "Projections for prevalence of Parkinson" disease and its driving factors in 195 countries and territories to 2050: modelling study of Global Burden of Disease Study 2021," *BMJ*, Jg. 388, 2025.

[2] C. Eggers *et al.,* "Versorgung von Parkinson-Patienten in Deutschland: Status quo und Perspektiven im Spiegel des digitalen Wandels," *Der Nervenarzt*, Jg. 92, Nr. 6, S. 602–610, 2021.

[3] C.-A. Holst, F. Wiegräbe, C. Redecker und V. Lohweg, "Die Parkinson-Erkrankung im Zeitalter der KI: Eine Smartphone-App als Schlüssel zur nutzerzentrierten Patientenversorgung," in *Künstliche Intelligenz im Einsatz für die erfolgreiche Patientenreise: Innovation, Integration und Stärkung*, M. A. Pfannstiel, Hg., Wiesbaden: Springer Gabler, 2025, S. 191–218.

[4] R. Gong *et al.,* "A survey of low-bit large language models: Basics, systems, and algorithms," *Neural Networks*, Jg. 192, S. 107856, 2025.

[5] Y. Gao *et al.,* "Retrieval-Augmented Generation for Large Language Models: A Survey," *arXiv:*2312.10997, 2024.

[6] A. J. Thirunavukarasu *et al.,* "Large language models in medicine," *Nature Medicine*, Jg. 29, Nr. 8, S. 1930–1940, 2023.

[7] Y. Kim *et al.,* "MDAgents: An Adaptive Collaboration of LLMs for Medical Decision-Making," in *Advances in Neural Information Processing Systems*, A. Globerson et al., Hg., Bd. 37, 2024, S. 79410–79452.

[8] M. Rehman *et al.,* "Advancement in medical report generation: current practices, challenges, and future directions," *Medical & Biological Engineering & Computing*, Jg. 63, Nr. 5, S. 1249–1270, 2025.

[9] Y. Huang *et al.,* "TrustLLM: Trustworthiness in Large Language Models," *arXiv:*2401.05561, 2024.

[10] X. Zhang, C. Tian, X. Yang, L. Chen, Z. Li und L. R. Petzold, "AlpaCare: Instruction-tuned Large Language Models for Medical Application," *arXiv:*2310.14558, 2025.

[11] L. Cardenas, K. Parajes, M. Zhu und S. Zhai, "AutoHealth: Advanced LLM-Empowered Wearable Personalized Medical Butler for Parkinson's Disease Management," in *2024 IEEE 14th Annual Computing and Communication Workshop and Conference (CCWC)*, 2024, S. 375–379.

[12] T. Dettmers, M. Lewis, Y. Belkada und L. Zettlemoyer, "GPT3.int8: 8-bit Matrix Multiplication for Transformers at Scale," in *Advances in Neural Information Processing Systems*, S. Koyejo et al., Hg., Bd. 35, 2022, S. 30318–30332.

[13] C. Zhang, J. Cheng, G. A. Constantinides und Y. Zhao, "LQER: low-rank quantization error reconstruction for LLMs," in *Proceedings of the 41st International Conference on Machine Learning*, 2024.

[14] G. Henry, P. T. P. Tang und A. Heinecke, "Leveraging the bfloat16 Artificial Intelligence Datatype For Higher-Precision Computations," in *2019 IEEE 26th Symposium on Computer Arithmetic (ARITH)*, 2019, S. 69–76.

[15] T. Dettmers, A. Pagnoni, A. Holtzman und L. Zettlemoyer, "QLoRA: Efficient Finetuning of Quantized LLMs," in *Advances in Neural Information Processing Systems*, A. Oh et al., Hg., Bd. 36, 2023, S. 10088–10115.

[16] Q. Jin *et al.,* "PubMedQA: A Dataset for Biomedical Research Question Answering," in *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, K. Inui et al., Hg., 2019, S. 2567–2577.

[17] T. Zhang *et al.,* "BERTScore: Evaluating Text Generation with BERT," in *International Conference on Learning Representations*, 2020.

DOI 10.5162/iCCC2026/P30

Möbius-Shapley: Native Feature Attribution for Quantum Logic Decision Trees

Samar Dhahbi
Chair of Database and Information Systems
Brandenburg University of Technology
Cottbus, Germany
dhahbsam@b-tu.de

Ingo Schmitt
Chair of Database and Information Systems
Brandenburg University of Technology
Cottbus, Germany
schmitt@b-tu.de

***Abstract*—The demand for transparent AI has made explainability crucial, particularly in high-stakes domains like healthcare and finance. Quantum Logic Decision Trees (QLDTs) offer interpretable classification but remain challenging for non-experts of logic to understand. Methods like SHAP provide model-agnostic explanations. However, there is a mismatch between the minterm-based QLDT's logic of feature negation and SHAP's assumption of feature missingness. We propose Möbius-Shapley, a native local (i.e., instance-specific) explanation method that bridges this gap by applying the Shapley value framework directly to Möbius-transformed QLDT weights. Our approach eliminates the semantic mismatch and provides native explanations, though we note that the exact derivation is computationally intensive for high-dimensional datasets compared to approximation methods.**

***Index Terms*—Explainable AI, Quantum Logic Decision Trees, Shapley Values, Möbius Transform, Feature Attribution**

I. Introduction

SHAP [1] (SHapley Additive exPlanations) has emerged as one of the most popular methods in explainable AI. Its game-theoretic framework for feature attribution bears structural similarity to Quantum Logic Decision Trees [2] (QLDTs). Both analyze feature combinations. However, a critical semantic mismatch exists: standard SHAP treats QLDTs as black boxes, approximating feature contributions by marginalizing over "absent" features. This contradicts the underlying minterm structure of QLDTs, where features are never missing but negated by the means of quantum logic.

This mismatch motivates our key insight: rather than applying generic SHAP, we can directly compute Shapley values from QLDT's native minterm representation. We introduce Möbius-Shapley, which transforms minterm weights to Möbius weights in order to avoid negation. This transformation enables faithful Shapley-based local (i.e., instance-specific) explanations tailored to QLDT's quantum logic semantics. Our approach resolves SHAP's "cancellation effect" of conflicting interactions and prevents the "impossible data" problem arising from standard marginalization. Möbius-Shapley approach shift in methodology fundamentally alters the nature of the explanation: whereas standard SHAP asks: "What makes this prediction different from the average case?" (a comparative analysis against a background), Möbius-Shapley asks: "How is this specific decision derived?" (a constructive analysis of a logic expression).

II. Literature Review

A. QLDT Fundamentals

A QLDT [2] is a two-class classifier based on CQQL [3], a quantum logic approach. In contrast to fuzzy logic [4], in CQQL all laws of a Boolean algebra are respected. Its evaluation is based on sums and products on n normalized attribute values from the unit interval $[0, 1]$.

1) Logic Representation: The QLDT tree is a visual representation of a logic-based model, it is constructed from a CQQL [3] expression e in disjunctive normal form (DNF), which is a disjunction of minterms. Each minterm is a conjunction of all n atomic conditions c_j, where c_j is a unary condition on a value of an attribute j that appears either in its positive (c_j) or negated ($\neg c_j$) form. Since each condition has two possible states, there are 2^n possible minterms. For example, for two attributes where `age` stands for high age and `income` stands for high income, we have four minterms: ($\neg$`age` $\wedge$ $\neg$`income`), (`age` $\wedge$ $\neg$`income`), ($\neg$`age` $\wedge$ `income`), (`age` $\wedge$ `income`). Every logical expression e can be represented by use of minterm weights $\theta_i \in \{0, 1\}$ indicating whether a minterm i is active (`true` $\leftrightarrow$ 1) or not (`false` $\leftrightarrow$ 0):

$$e = \bigvee_{i=0}^{2^n-1} (minterm_i \wedge \theta_i)\,. \tag{1}$$

From a logic expression e, a QLDT ($qldt(e)$) is generated which represents a compact visualization of e in form of a kind of a decision tree. See Figure 1.

2) Evaluation: Unlike classical trees that follow one path down to a single leaf, the evaluation of a QLDT involves navigating from the root to all active leaves in parallel.

Let an object $o = \{a_1^o, a_2^o, \ldots, a_n^o\}$ where $a_j^o \in [0, 1]$ is the value of the j-th attribute a_j of o, and $[\cdot]^o \in [0, 1]$ denote the evaluation of a CQQL condition on o:

$$[minterm_i]^o = \prod_{j=1}^{n} [c_{ij}]^o. \tag{2}$$

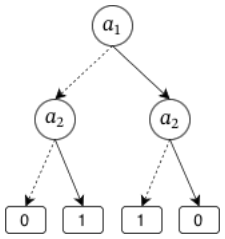
where c_{ij} refers to the j-th attribute condition c_j of $minterm_i$, if c_{ij} is positive then $[c_{ij}]^o := a_j^o$ otherwise $[c_{ij}]^o := 1 - a_j^o$.

Like the example shown in Figure 1, every minterm represents a path from the root to a leaf, where the right solid arrow indicates a positive condition and the left dashed arrow indicates its negation.

The results from all class-1-leaves are summed up to produce a class value in $[0, 1]$:

$$[e]^o = \sum_{i=0}^{2^n-1} \theta_i \prod_{j=1}^{n} [c_{ij}]^o =: [qldt(e)]^o. \quad (3)$$

A final threshold τ is applied to the result value from (3) for obtaining a discrete class decision (0 or 1).



$$e = (a_1 \wedge \neg a_2) \vee (\neg a_1 \wedge a_2)$$
$$[e]^o = [qldt(e)]^o = a_1^o \cdot (1 - a_2^o) + (1 - a_1^o) \cdot a_2^o$$

Fig. 1. A QLDT for an XOR example

B. SHAP

SHAP [1] (SHapley Additive exPlanations) adapts Shapley values [5] from cooperative game theory to machine learning explainability. In this framework, a model prediction is viewed as a game where the input features are the "players," and the specific prediction value is the "payout". For a set N of n features, the goal is to distribute the total "payout" (the difference between the model's prediction for a specific instance and the average prediction across the dataset) fairly among the features. The contribution of a feature j, denoted as φ_j, is calculated as the weighted average of its marginal contributions across all possible feature coalitions:

$$\varphi_j = \sum_{S \subseteq N \setminus \{j\}} \frac{|S|!(n - |S| - 1)!}{n!} [v(S \cup \{j\}) - v(S)]. \quad (4)$$

where the core component of this calculation is the characteristic set function $v(S)$, which defines the model's output when a subset of features S is present. In the context of machine learning, features not in S (i.e., $N \setminus S$) are considered "missing". Since most models cannot handle missing inputs natively, SHAP approximates $v(S)$ via marginalization. It integrates over the missing features using a background dataset or an assumed distribution. SHAP holds the efficiency property, which guarantees that the sum of the feature attributions equals the total deviation of the prediction from the baseline:

$$v(N) - v(\varnothing) = \sum_j \varphi_j. \quad (5)$$

where $v(\varnothing)$ is the average prediction of background data.

III. Motivation

The structural similarity, combined with the critical semantic gap between minterms and the player set motivated our work to develop a native Shapley-based method for QLDTs without treating them as black boxes as it is done in the SHAP approach. By solely relying on the instance's native minterm representation rather than on average contribution, our method operates independently of background data distributions. Consequently, this leads to more transparent interpretation through a constructive breakdown of the specific decision.

A. The Structural Similarity

Interestingly, QLDTs inherently represent a form of coalitions, where each minterm evaluates a specific combination of feature presences and negations. This structure bears remarkable resemblance to the Shapley value framework, where the prediction-baseline difference (5) is decomposed into feature contributions.

B. Structure Incompatibility

Despite the similarity, there is a fundamental semantic gap. In minterms of quantum logic, all features are inherently present (positive or negated) with continuous truth values. The concept of "missing" or "absent" features in SHAP contradicts the foundational principles of minterms of a QLDT. In SHAP, when a feature is excluded from a coalition, it is treated as "missing" and marginalized over a background dataset. In minterms of a QLDT, features are never missing, negation $(1 - x)$ explicitly represents logical complement.

This gap means that while both frameworks compute contributions to feature combinations, they interpret "feature absence" differently. SHAP's marginalization approach may induce artificial scenarios. For example, in an exclusive-OR (XOR) rule, marginalizing over a missing feature can make it seem like the feature has no effect on average. This misleading cancellation can obscure the feature's critical role for a specific input. This weak point will be discussed in Section V.

IV. Methodology: The Möbius-Shapley Approach

We propose Möbius-Shapley, a native explanation method that bridges the gap between QLDT logic and Shapley analysis by utilizing the Möbius transformation.

A. Core Idea

Our approach leverages the unique structure of QLDTs to provide native explanations rather than treating the model as a black box. Our key insight recognizes that the Möbius transformation [6] can convert QLDT's minterm-based evaluation into a form compatible with Shapley analysis. By transforming minterms into feature interactions, we eliminate the negation/missingness mismatch while preserving all logical relationships.

B. Möbius Transformation

This mathematical transformation serves as the crucial translator between QLDT logic and Shapley analysis.

In our context, a minterm weight is interpreted as a value of set function $\theta_i = m(X)$ where $X \subseteq N$ is the set of features that appear positively in minterm i. Möbius transformation, like discussed in [7], maps $m(X)$ into interaction weights $\mu(X)$:

$$\mu(X) = \sum_{Y \subseteq X} (-1)^{|X|-|Y|} \cdot m(Y). \tag{6}$$

Let $\lambda_X(o)$ be the product of the feature values in the set X of an object o. Möbius transform maps the evaluation $[e]^o = \sum_{i=0}^{2^n-1} \theta_i \cdot [minterm_i]^o$ to an equivalent negation-free representation:

$$[e]^o = \sum_{X \subseteq N} \mu(X) \cdot \lambda_X(o). \tag{7}$$

For example, for an object $o = \{a_1, a_2\}$ instead of evaluating the expression $e = (a_1 \wedge \neg a_2)$ to $a_1 \cdot (1 - a_2)$, we compute it with a negation-free form as $[e]^o = \mu(\varnothing) \cdot 1 + \mu(\{a_1\}) \cdot a_1 + \mu(\{a_2\}) \cdot a_2 + \mu(\{a_1, a_2\}) \cdot a_1 \cdot a_2$.

For the expression $(a_1 \wedge \neg a_2)$, the Möbius weights computed from (6) yields:

$$\mu(\varnothing) = 0, \quad \mu(\{a_1\}) = 1, \quad \mu(\{a_2\}) = 0, \quad \mu(\{a_1, a_2\}) = -1$$

Substituting these values yields $[e]^o = 0 \cdot 1 + 1 \cdot a_1 + 0 \cdot a_2 + (-1) \cdot a_1 \cdot a_2 = a_1 - a_2 \cdot a_1$ which matches the original evaluation.

C. Method Overview

Our QLDT native feature attribution algorithm leverages the algebraic properties of the Möbius transform to solve the feature absence problem in QLDTs. The core innovation of this method lies in its formulation of the coalition value function $v(S)$ in (4). For a given instance o and a feature subset S, the value $v_o^{m\ddot{o}}(S)$ is defined as:

$$v_o^{m\ddot{o}}(S) = \sum_{X \subseteq S} \mu(X) \cdot \lambda_X(o). \tag{8}$$

This formulation replaces SHAP's marginalization approach, allowing us to compute contributions exactly from the model's interaction weights. Applying the Shapley formula in (4), we get the Möbius-Shapley values $\varphi_j^{m\ddot{o}}$ for a feature j:

$$\varphi_j^{m\ddot{o}} = \sum_{S \subseteq N \setminus \{j\}} \frac{|S|!(n - |S| - 1)!}{n!} [v_o^{m\ddot{o}}(S \cup \{j\}) - v_o^{m\ddot{o}}(S)]. \tag{9}$$

Our approach offers advantages over SHAP: it requires no background data, as the Möbius weights hold all necessary information. It also makes exact computations, eliminating the need for sampling approximation. Additionally, Möbius-Shapley method satisfies the efficiency axiom (5) where:

$$v_o^{m\ddot{o}}(N) - v_o^{m\ddot{o}}(\varnothing) = \sum_j \varphi_j^{m\ddot{o}}. \tag{10}$$

Based on (7) and (8), the full coalition value $v_o^{m\ddot{o}}(N)$ corresponds to the model evaluation $[e]^o$. For the empty set, equations (7) and (6) show that $v_o^{m\ddot{o}}(\varnothing) = \mu(\varnothing) = m(\varnothing) \in \{0, 1\}$. Since this term is instance-independent, we denote the constant baseline as $v^{m\ddot{o}}(\varnothing)$ thus $[e]^o = \sum_j \varphi_j^{m\ddot{o}} + v^{m\ddot{o}}(\varnothing)$.

The definition (10) marks a fundamental shift in baselines: While both methods utilize a constant reference point, standard SHAP derives its baseline **statistically** from the background dataset (the average prediction). In contrast, Möbius-Shapley derives its baseline **structurally** from the model itself (the weight of the all-negated minterm). Thus, SHAP measures deviation from the 'average case,' while Möbius-Shapley measures construction from the logical 'ground state'.

V. Results and Discussion

To validate the Möbius-Shapley framework, we applied the method to a QLDT modeling of an exclusive-OR (XOR) problem. This section details the feature attribution process and analyzes the advantages of this native approach compared to standard model-agnostic methods like SHAP.

A. A Real-World Example

We consider a cheese quality assessment based on storage conditions, modeled by two normalized features: `humidity` (H) and `temperature` (T). The model classifies a cheese as high quality (class 1) if the storage conditions are contrasting (e.g., high `humidity` with low `temperature`, or low `humidity` with high `temperature`). Conversely, it is classified as low quality (class 0) if both conditions are similarly high or low. This is modeled by a QLDT expression $e = (\neg \mathtt{H} \wedge \mathtt{T}) \vee (\mathtt{H} \wedge \neg \mathtt{T})$ (XOR) and a threshold $\tau = 0.55$. For a specific instance o where the cheese ring got stored in a cave with high `humidity` value ($\mathtt{H} = 0.7$) and low temperature ($\mathtt{T} = 0.3$), the QLDT evaluation is $[e]^o = (1 - 0.7) \cdot 0.3 + 0.7 \cdot (1 - 0.3) = 0.58$. Since its evaluation is higher than τ, this cheese ring is predicted as high quality (class 1). To explain this instance using Möbius-Shapley, first, we transform the QLDT minterm weights into interaction weights (μ) using Equation (6). This transformation reveals the feature interaction:

$$\begin{aligned}
\mu(\varnothing) &= m(\varnothing) = 0 \\
\mu(\{\mathtt{T}\}) &= m(\{\mathtt{T}\}) - m(\varnothing) = 1 - 0 = 1 \\
\mu(\{\mathtt{H}\}) &= m(\{\mathtt{H}\}) - m(\varnothing) = 1 - 0 = 1 \\
\mu(\{\mathtt{H}, \mathtt{T}\}) &= m(\{\mathtt{H}, \mathtt{T}\}) - m(\{\mathtt{H}\}) - m(\{\mathtt{T}\}) + m(\varnothing) \\
&= 0 - 1 - 1 + 0 = -2
\end{aligned}$$

Using the Möbius weights, we calculate the coalitional values according to Equation (8):

$$\begin{aligned}
v_o^{m\ddot{o}}(\varnothing) &= \mu(\varnothing) \cdot 1 = 0 \\
v_o^{m\ddot{o}}(\{\mathtt{H}\}) &= \mu(\varnothing) \cdot 1 + \mu(\{\mathtt{H}\}) \cdot \mathtt{H} = 0 + 1 \cdot 0.7 = 0.7 \\
v_o^{m\ddot{o}}(\{\mathtt{T}\}) &= \mu(\varnothing) \cdot 1 + \mu(\{\mathtt{T}\}) \cdot \mathtt{T} = 0 + 1 \cdot 0.3 = 0.3 \\
v_o^{m\ddot{o}}(\{\mathtt{H}, \mathtt{T}\}) &= \mu(\varnothing) + \mu(\{\mathtt{H}\}) \cdot \mathtt{H} + \mu(\{\mathtt{T}\}) \cdot \mathtt{T} \\
&\quad + \mu(\{\mathtt{H}, \mathtt{T}\}) \cdot \mathtt{H} \cdot \mathtt{T} \\
&= 0 + 0.7 + 0.3 + (-2) \cdot 0.21 = 0.58 = [e]^o
\end{aligned}$$

After applying Shapley formula (9), we obtain $\varphi_{\mathtt{H}}^{m\ddot{o}} = 0.49$ and $\varphi_{\mathtt{T}}^{m\ddot{o}} = 0.09$, which means that `humidity` contributes $+0.49$ of the total evaluation and `temperature` contributes $+0.09$.

B. Interpretation of Results

The results derived in the previous subsection align with the underlying XOR logic of the model in two ways:

1. Exclusivity: The calculated Möbius interaction weight of $\mu(\{\mathtt{H}, \mathtt{T}\}) = -2$ serves as a mathematical penalty. This negative value explicitly aligns with the XOR function, reducing the prediction value when both features are present simultaneously.

2. Dominance: For the specific input $(\mathtt{H} = 0.7, \mathtt{T} = 0.3)$ the condition $(\mathtt{H} \wedge \neg\mathtt{T})$ is the dominant active path. Feature `H` receives a higher attribution ($\varphi_{\mathtt{H}}^{m\ddot{o}} = 0.49$) because it strongly satisfies the positive condition, whereas Feature `T` contributes positively ($\varphi_{\mathtt{T}}^{m\ddot{o}} = 0.09$) but having a low value, effectively satisfying the negation $\neg\mathtt{T}$.

C. Comparative Analysis: Möbius-Shapley vs. Standard SHAP

While standard SHAP (e.g., KernelSHAP [1]) is effective for black box models, our analysis highlights that the native Möbius-Shapley approach offers distinct advantages by exploiting the QLDT structure. It eliminates marginalization, which prevents the "cancellation effect" and ensures the explanation is not derived from logically impossible data combinations.

1) Avoiding The Cancellation Effect: A critical limitation of standard SHAP is its tendency to obscure conflicting interactions by averaging them into a single attribution value. X. Huang and J. Marques-Silva [8] argue that marginalization can cause these positive and negative effects to cancel out, potentially resulting in near-zero Shapley values that imply feature irrelevance. For example, in the XOR scenario described in Subsection V-A, the features `humidity` and `temperature` are individually necessary but mutually exclusive. Regarding `humidity`: in a "high `temperature`" context, a storage cave being not humid might have a positive impact (+1), while in a "low `temperature`" context, it might be negative (-1). Standard SHAP averages these (0 contribution), obscuring the feature's importance. Möbius-Shapley avoids this by utilizing explicit interaction weights $\mu(\{\mathtt{H}, \mathtt{T}\}) = -2$, that act like a penalty when both features are present, providing a more "white-box" explanation.

2) Handling Feature Dependencies: The "Impossible Data" Problem: Standard SHAP methods typically assume feature independence when marginalizing over absent features. As Aas et al. [9] demonstrate, this assumption forces the model to average over impossible data combinations that don't make sense in real life. For example, when explaining a model that uses the features `sex` and `pregnant`, SHAP might create invalid data (e.g., `male` AND `pregnant`) to compute marginal contributions, which can lead to inaccurate explanations. In contrast, the Möbius-Shapley approach is native to the QLDT; it computes contributions solely from minterms weights, inherent in the model structure. By utilizing the explicit interaction weights (μ) derived from the Möbius transform, our method respects the Boolean dependencies without approximating missing features via unrealistic sampling.

3) Contrastive vs. Constructive Explanations: Both SHAP and Möbius-Shapley have inherent limitations. While SHAP's model-agnostic nature provides broad applicability, its reliance on marginalization can fail in QLDTs by causing semantic mismatches, cancellation effects, and impossible data scenarios. Möbius-Shapley attempts to resolve these issues by leveraging the native minterm representation of QLDTs, but it is computationally expensive due to the exponential number of minterms (2^n). While both methods employ Shapley values, they end up computing different values. Applying SHAP to the same example from Subsection V-A, we got: $\varphi_{\mathtt{H}} = 0.04$ and $\varphi_{\mathtt{T}} = 0.09$ with a base value $v(\varnothing) = 0.45$. Thus, they answer fundamentally different interpretative questions. Standard SHAP provides **contrastive explanations** that ask: "how does this prediction differ from the average?" by measuring feature contributions relative to the population baseline $v(\varnothing)$. In contrast, Möbius-Shapley provides **constructive explanations** that ask: "how is this specific prediction built?" by decomposing the actual evaluation starting from the all-negated baseline $v^{m\ddot{o}}(\varnothing)$. This distinction explains why the methods yield to different attribution values; they measure different quantities.

In our example: the cheese ring gets 0.58 quality score (prediction: $1 \leftrightarrow$ high quality).

TABLE I
SHAP VS. MÖBIUS-SHAPLEY EXPLANATION COMPARISON

Method	SHAP	Möbius-Shapley
Baseline	0.45 (global average)	0 (ground state)
`humidity`	+0.04	+0.49
`temperature`	+0.09	+0.09
Explanation	This cheese is high quality due to storage being warmer and slightly more humid than average	`humidity` is the main driver, with `temperature` providing secondary support

The question that arises is: If SHAP and Möbius-Shapley provide different values, then which of them are better for interpreting?

The answer is that both methods have complementary purposes. SHAP can be used to identify what makes a case unusual, while Möbius-Shapley can be used to understand how the model constructs the decision, revealing the internal reasoning and feature interactions.

VI. CONCLUSION

We have presented Möbius-Shapley, a native feature attribution method for Quantum Logic Decision Trees that resolves the fundamental semantic conflict between SHAP's "missingness" and QLDT's "negation". By leveraging the Möbius transformation, our approach derives non-approximated explanations directly from the model's structure, eliminating the

need for background data sampling and effectively addressing the “cancellation problem” in conflicting interactions. However, both approaches compute different contribution values, meaning that they answer different questions. SHAP is used for contrastive explanation whereas our approach provides constructive explanations.

Future work will focus on optimizing the computational efficiency of the Möbius transformation, particularly for high-dimensional datasets. We plan to investigate the application of this method to pruned QLDT to enhance scalability while maintaining interpretability.

REFERENCES

[1] S. M. Lundberg and S.-I. Lee: A Unified Approach to Interpreting Model Predictions, *Advances in Neural Information Processing Systems*, vol. 30, pp. 4765–4774, 2017.

[2] I. Schmitt, ”QLDT: A Decision Tree Based on Quantum Logic”, in *New Trends in Database and Information Systems – ADBIS 2022 Short Papers, Doctoral Consortium and Workshops*, 2022.

[3] I. Schmitt, ”QQL: A DB&IR Query Language”, *The VLDB Journal*, vol. 17, no. 1, pp. 39–56, 2008.

[4] L.mZadeh, ”Fuzzy sets”, *Journal of Information and Control*, 8, 1965.

[5] L. S. Shapley, ”A Value for n-Person Games”, *Contributions to the Theory of Games*, vol. 2, no. 28, pp. 307–317, 1953.

[6] G.-C. Rota, ”On the Foundations of Combinatorial Theory: I. Theory of Möbius Functions”, in *Classic Papers in Combinatorics*, pp. 332–360, Springer, 1964.

[7] G. Wirsching, I. Schmitt, and M. Wolff, ”Ausgewählte Anwendungen”, in *Quantenlogik Band 1*, Springer Vieweg, 2025.

[8] X. Huang and J. Marques-Silva, ”On the failings of Shapley values for explainability,” , in *International Journal of Approximate Reasoning*, , vol. 171, Art. no. 109112, Jan. 2024,

[9] K. Aas, M. Jullum, and A. Løland, ”Explaining Individual Predictions When Features Are Dependent: More Accurate Approximations to Shapley Values”, *Artificial Intelligence*, vol. 298, p. 103502, 2021.

DOI 10.5162/iCCC2026/P31

A Cybersecurity Approach for Visual Data Flow in Digital Twins: A Real-World Case Study

Hamid Zargariasl, Aghaali Faraj, Christian Borck, Martin Behm, Annamria Arnouk, Christian Herglotz,
Brandenburg University of Technology, Cottbus, Germany
zargaria@b-tu.de

Abstract

We considered a digital twin system designed to handle visual information through virtual reality (VR) and fixed cameras. This system is part of a digital twin framework that has been proposed and managed over several years, specifically related to the production line of a component of an airplane engine. Several relevant tools are integrated into this framework using technologies such as HTTP, gRPC and WebRTC. Our approach involves analysing the topology of the digital twin system and conducting a threat analysis using the STRIDE model. Based on this analysis, we propose mitigation strategies to enhance the resilience of the current configuration against potential cyber-attacks.

1 Introduction

Digital twins are generally defined as virtual representations of physical assets, systems, or processes that are continuously updated with real-time data, enabling analysis, prediction, and optimization across the asset's lifecycle. A digital twin integrates the physical and virtual spaces through data, models, and services, forming a closed loop of information that supports monitoring, simulation, and decision-making. The digital twin as a three-part structure composed of the physical entity, the virtual entity, and the data connections between them, emphasizing its role in lifecycle management and predictive analytics. Together, these definitions highlight the importance of real-time data integration and bidirectional communication as foundational characteristics of digital twin systems ([1],[2]).

Digital twins have gained significant attention in recent years as a key enabler for integrating physical systems with their virtual counterparts, enabling real-time monitoring, analysis, and optimization. Researchers increasingly emphasize the role of digital twins in improving system performance, enhancing predictive maintenance, and supporting data-driven decision-making across diverse domains such as manufacturing, transportation, and smart cities. These studies illustrate the expanding scientific focus on digital twins as an essential component of cyber-physical systems ([1],[2]).

Handling visual information within digital twin systems poses significant challenges due to the large volume, high dimensionality, and real-time processing requirements of such data. Beyond the inherent computational complexity, these visual pipelines introduce additional vulnerabilities that must be addressed through robust cybersecurity approaches. As a result, ensuring secure and efficient processing of visual information has become a critical aspect of maintaining the reliability and resilience of digital twin environments.

This paper investigates a real-world environment for implementing a digital twin system, incorporating video information, where multiple components interact using diverse communication protocols. Our objective is to analyze the network topology to identify its most vulnerable points and to propose suitable mitigation methods against potential cyberattacks.

Chapter 2 provides an overview of the digital twin scenario under study. Chapter 3 details the network analysis methodology and the rationale behind the proposed mitigation strategies. Finally, the conclusion summarizes the key findings and discusses future research directions.

2 A Real World Digital Twin

We utilize a camera-based topology during the assembly of an engine component, as previously reported [3]. The digital twin is implemented using the Asset Administration Shell (AAS), which provides a standardized framework for representing and managing the asset's data and functions. AAS is a standardized digital representation of an asset in Industry 4.0. It defines how data, functions, and services related to an asset are structured and accessed. The hardware topology used in this process is depicted in Figure 1.

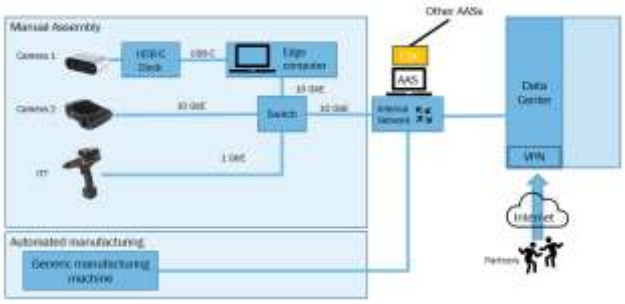


Figure 1 The topology of connected hardware including digital twin (AAS node).

The visual system, comprising augmented reality (AR) and cameras, is capable of detecting any aspect of the work and automatically notifying the worker. This visual information is integral to the digital twin aspect of the system. Various tools, such as AFOS (Agent and Server), CTDS (Tacit Data Server), CTDS (Digital Assembly Master), Nucleus Server, NVIDIA 3D Asset Bridging, and others, are interconnected via HTTP, gRPC, and WebRTC. However,

DOI 10.5162/iCCC2026/P31

delve into those specific details here as our primary focus is on cybersecurity issues, which are addressed in Figure 2.

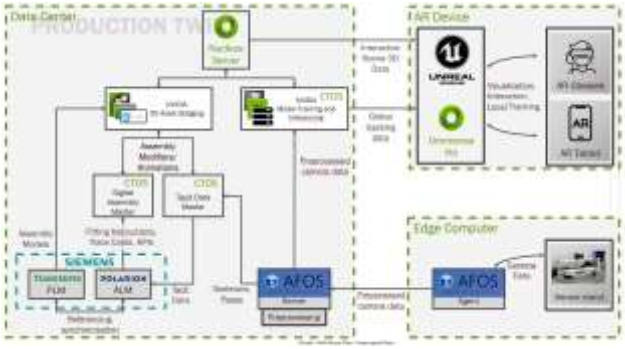


Figure 2 The topology of connected tools.

3 Threat Analysis

The STRIDE threat analysis model is one of the most established frameworks for systematically identifying security threats in distributed and networked systems. Originally introduced by Microsoft, STRIDE categorizes threats into six classes—Spoofing, Tampering, Repudiation, Information Disclosure, Denial of Service, and Elevation of Privilege—allowing researchers to perform structured evaluations of system vulnerabilities. Its applicability has been demonstrated across various cyber-physical systems, including IoT, cloud infrastructures, and digital twins ([4-6]). These references establish STRIDE as a robust and academically validated approach for threat identification in complex digital environments.

The STRIDE threat modeling framework is aligned with the requirements and terminology of ISO/IEC 27005 and ISO/IEC 15408, which provide standardized methodologies for information-security risk assessment and evaluation. We applied the STRIDE threat-modeling framework, and the resulting analysis is summarized in Table 1.

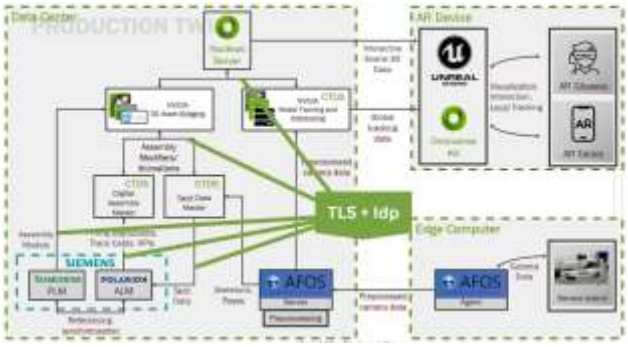
STRIDE threat	Where it occurs	Mitigation in design
S-Spoofing	gRPC (AFOS aganet and Server)	mTLS with CA issued certificates
S-Spoofing	Nucleus, APIe, USD Connectors	idP issued JWT tokens +TLS
S-Spoofing	AR/WebRTC Signals	idP Authentication+DTLS
T-Tampering	gRPC Traffic	TLS integrity through mTTS
T-Tampering	HTTP (Nucleus /APIs/Connectors)	TLS 1.2 /1.3
T-Tampering	WebRTC media /data	SRTP +DTLS
R- Repudiation	Backend Servics	Certificate Identity +logging
R- Repudiation	Nucleus and APIs	JWT tokens+server side logs
R- Repudiation	AR/WebRTC Signals	Authentication signalling logs
I- Information Disclosure	gRPC data	mTLS encryption
I- Information Disclosure	Nucleus/APIs/USD Connectors	TLS encryption
I- Information Disclosure	WebRTC AR Streams	DTLS +SRTP
D- Denial of Service	Backend, APIs/ Signals, WebRTC	Certificate and token checkreduce unauthorized traffic
E- Elevation of Privilage	Nucleus and APIs	idP role claims+ACL enforcement
E- Elevation of Privilage	Backend Ervices	mTLS identity binding

Table 1 STRIDE threats and mitigations.

We focused more on the communication links between the software components in our system and identified three principal communication methods as a use case: gRPC, HTTP, and WebRTC. These technologies were selected due to their widespread adoption and strong scientific foundation, each of which is well documented in existing literature. Specifically, recent studies demonstrate the efficiency and scalability of gRPC in distributed systems, the robustness and interoperability of HTTP in heterogeneous environments, and the low-latency, real-time capability of WebRTC for peer-to-peer communication [7-10]. Together, these publications provide a solid scientific basis for understanding and evaluating the communication mechanisms used in our environment.

We systematically designed mitigation measures adapted to the security characteristics of each individual communication protocol:

DOI 10.5162/iCCC2026/P31



A**. mTLS for gRPC**

Mutual TLS (mTLS) is an extension of standard Transport Layer Security in which both the client and the server authenticate each other using X.509 certificates. This bidirectional authentication ensures that only trusted components can join a distributed system, which is particularly important in microservice-based architectures using gRPC. Because gRPC is built on HTTP/2 and is commonly used for high-performance inter-service communication, securing these channels with mTLS provides confidentiality, integrity, and strong identity binding for each endpoint. The use of mTLS is widely recommended in cloud-native environments due to its proven security properties and compatibility with service meshes such as Envoy or Istio. The security foundation of mTLS is defined in the TLS 1.3 standard (RFC 8446) [7]. Its usage in our network is illustrated in Figure 3.

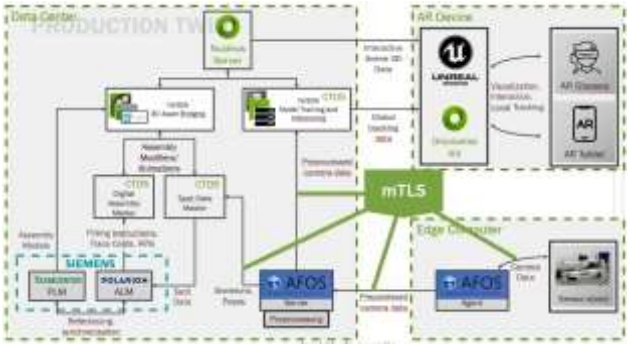


Figure 3 mTLS for gRPC.

B. **TLS + Identity Provider (IdP) for HTTP**

For HTTP-based communication, a widely adopted security pattern combines TLS encryption with authentication and authorization managed by an external Identity Provider (IdP). TLS provides transport-layer confidentiality and integrity for data exchanged between clients and web services, whereas the IdP—commonly implementing OAuth 2.0 or OpenID Connect—issues verifiable tokens representing user or service identities. This separation of concerns enables secure scalability across distributed digital-twin platforms and enterprise service architectures. TLS 1.3 (RFC 8446) establishes the cryptographic foundation for secure transmission [7], while the identity-federation mechanisms are specified in "The OAuth 2.0 Authorization Framework" (RFC 6749) [8]. The integration of these technologies offers robust authentication, flexible authorization, and encrypted transport without requiring modifications to the HTTP application logic. The usage of this components in our network is illustrated in Figure 4.

Figure 4 TLS + Identity Provider (IdP) for HTTP.

C. **DTLS + SRTP for WebRTC**

WebRTC real-time communication relies on Datagram Transport Layer Security (DTLS) and Secure Real-time Transport Protocol (SRTP) to protect media and data streams over peer-to-peer connections. DTLS provides TLS-equivalent security over UDP, enabling certificate-based authentication and encryption during the handshake phase. Following key exchange, SRTP encrypts and authenticates audio and video packets efficiently, minimizing latency—an essential requirement for AR/VR and interactive real-time systems. DTLS is formally specified in RFC 9147 [9], while SRTP is standardized in RFC 3711 [10]. Together, DTLS and SRTP constitute the mandatory security architecture of WebRTC, ensuring confidentiality, integrity, replay protection, and secure key negotiation for real-time multimedia communication. Figure 5 illustrates their usage in our network.

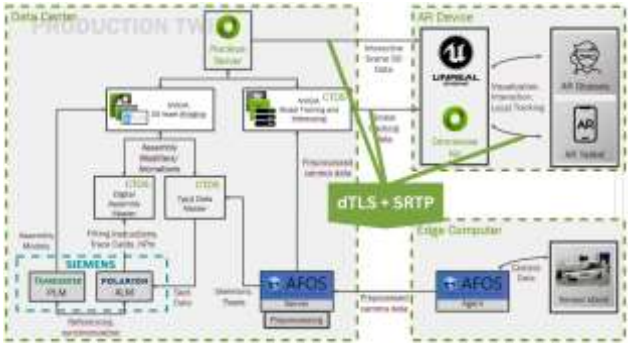


Figure 5 DTLS + SRTP for WebRTC.

4 Conclusion

In this study, we examined a real-world implementation of a digital twin, particularly focusing on the component handling video information. Our analysis included the evaluation of all software tools used and their connection protocols, employing the STRIDE threat analysis methodology. From this, we compiled a comprehensive list of necessary mitigation measures applicable to this environment. The digital twin in question is actively utilized for visual information management within the production line of specific types of airplane engines.

Our findings underscore the significance and complexity of cybersecurity challenges in this domain, exacerbated by the intricate nature of the software and communication protocols involved. As these complexities continue to escalate, we assert that cybersecurity design methods require heightened attention to effectively anticipate and mitigate future threats.

Future research could expand on this work by investigating larger and more complex implementations of digital twins,

further contributing to the development of robust cybersecurity strategies in increasingly sophisticated technological environments.

Acknowledgement

The project DIREKT and this publication were funded by the Federal Ministry for Economic Affairs and Climate Action (BMWK) on the basis of a decision by the German Bundestag (funding reference: 20L2108C1).

5 Reference

[1] M. Grieves and J. Vickers, “Digital Twin: Mitigating unpredictable, undesirable emergent behavior in complex systems,” in Transdisciplinary Perspectives on Complex Systems, F.-J. Kahlen, S. Flumerfelt, and A. Alves, Eds. Cham, Switzerland: Springer, 2017, pp. 85–113.

[2] F. Tao, H. Zhang, A. Liu, and A. Y. C. Nee, “Digital twin in industry: State-of-the-art,” IEEE Trans. Ind. Informatics, vol. 15, no. 4, pp. 2405–2415, Apr. 2019.

[3] H. Zargariasl, S. R. Khan, C. Borck, M. Behm, and C. Herglotz, “Threat analysis in real-world computer networks possessing a digital twin: Graph visualization and Markov clustering,” in Proc. IEEE Smart World Congress (SWC), Calgary, Canada, 2025.

[4] O. Saßnick, T. Rosenstatter, C. Schäfer, and S. Huber, “STRIDE-based methodologies for threat modeling of industrial control systems: A review,” in Proc. IEEE Int. Conf. on Industrial Cyber-Physical Systems (ICPS), 2024, pp. 1–8, doi: 10.1109/ICPS59941.2024.10639949.

[5] P. Das et al., “STRIDE-based cybersecurity threat modeling, risk assessment and treatment of an in-vehicle infotainment system,” Vehicles, vol. 6, no. 3, pp. 1140–1163, 2024.

[6] M. da Silva et al., “Automated ICS template for STRIDE Microsoft Threat Modeling Tool,” in Proc. 18th Int. Conf. on Availability, Reliability and Security (ARES), 2023.

[7] E. Rescorla, “The Transport Layer Security (TLS) protocol version 1.3,” RFC 8446, Aug. 2018. [Online]. Available: https://www.rfc-editor.org/rfc/rfc8446.

[8] D. Hardt, “The OAuth 2.0 authorization framework,” RFC 6749, Oct. 2012. [Online]. Available: https://www.rfc-editor.org/rfc/rfc6749

[9] E. Rescorla, N. Modadugu, and J. Hodges, “Datagram Transport Layer Security (DTLS) version 1.3,” RFC 9147, Apr. 2022. [Online]. Available: https://www.rfc-editor.org/rfc/rfc9147

[10] M. Baugher, D. McGrew, M. Naslund, E. Carrara, and K. Norrman, “The Secure Real-time Transport Protocol (SRTP),” RFC 3711, Mar. 2004. [Online]. Available: https://www.rfc-editor.org/rfc/rfc3711

DOI 10.5162/iCCC2026/P32

Visualizing Score Contributions in BBQ-Trees with Applications to Medical Screening Data

Alexander Stahl
FG Datenbank- und Informationssysteme
Brandenburgische Technische Universität
Cottbus, Germany
stahlale@b-tu.de

***Abstract*—Interpretable decisionmaking is essential in medical and health-related applications. BBQ-Trees (BBQTs) combine decision-tree structure with quantum-inspired logic, with advantages to accuracy and interpretability. However, their evaluation scheme can make per-sample explanations challenging. We propose a visualization method that maps subtree contributions to colors in order to reveal which decision paths drive an individual prediction. Using publicly available medical datasets as illustrative examples, we show that the method makes BBQT predictions more transparent without altering their predictive behavior.**

***Index Terms*—explainable AI, decision trees, visualization, clinical decision support, medical screening data**

I. Introduction

Machine learning (ML) is becoming increasingly popular in high-stakes domains such as medicine, where predictive models assist clinicians in diagnostic and therapeutic decision making. As summarized by Rahmani et al. [1], current developments show both the potential of ML in clinical practice but also the challenges associated with integrating complex models into safety-critical applications. While black-box models have gained widespread popularity and are increasingly deployed in many settings, recent works argue for the need for explainable models in high-stakes applications. As one of such applications, clinical decision support systems require models that are not only accurate but also interpretable. This is not only to foster trust among health practitioners and maintain accountability, but also to comply with regulatory standards.

Classical decision trees (DT) satisfy many of these requirements due to their transparent structure. However, they often struggle to represent curved or otherwise non-linear decision boundaries without becoming overly deep or fragmented, which can reduce interpretability [2], [3]. One alternative model, the *BBQ-Tree (BBQT)*, generalizes the DT and the so-called *Quantum Logic Decision Tree (QLDT)*. The model's QLDT components provide gradual, *quantum-inspired* truth values that allow the model to capture trend information and curved decision boundaries, whereas the DT components retain the capabilities of making classical axis-parallel attribute splits for accuracy boosts [4].

A inherited limitation of BBQTs is that all branches contribute to the final evaluation, due to the gradual truth semantics. Unlike classical DTs, where irrelevant branches are simply not traversed, BBQTs evaluate all decision paths. For bigger trees, this global evaluation can complicate the interpretability of the model and per-sample explanations, as it becomes more difficult to identify which parts of the tree were most influental for a particular prediction. In this work, we argue that local interpretability for BBQTs can be recovered by focusing on how the model evaluates a specific input sample. Instead of explaining all global interactions, we analyze the subtree scores that arise during the BBQT evaluation and visualize their relative contributions. Our contribution is a coloring method that maps subtree evaluation scores to a heat-based color scheme, which can highlight which parts of the BBQT contribute most to an individual prediction. This produces a locally interpretable representation of BBQT decisions, making the model's internal reasoning accessible to end users such as clinicians.

II. Related Work

Interpretable machine learning has produced a wide range of approaches for explaining model behavior, both for tree-based methods and for general black-box models. A substantial body of post-hoc explanation techniques aims to provide model-agnostic interpretability, while other work focuses on enhancing the transparency of tree structures themselves. Several tools and methods improve the visual interpretability of classical DTs. Common techniques include shading nodes by class distribution or prediction confidence, highlighting the path taken by an individual sample (e.g., `dtreeviz` [5]), and using leaf-level heatmaps to represent class purity, as implemented in `treeheatr` [6].

More general post-hoc methods such as LIME and SHAP approximate local feature importance by perturbation or additive attribution models. For neural architectures, techniques such as Grad-CAM and attention visualizations provide insight into salient regions or internal weight patterns. For an overview of existing methods, see Molnar [7]. While these approaches can highlight important features, they do not necessarily reveal the underlying *reasoning structure* of the model. In logic-based models, particularly tree-based ones, interpretability benefits from the ability to inspect the explicit decision structure, even when the tree becomes large or deep. This motivates staying within the family of tree-based models rather than relying solely on model-agnostic post-hoc explanations.

DOI 10.5162/iCCC2026/P32

Our method extends the idea of path- or node-level visualization by mapping subtree evaluation scores to color intensities. This provides a sample-specific explanation of BBQT predictions, which bridges the gap between the interpretability of logic-base models and post-hoc explanation techniques.

III. Background: BBQT and QLDT

This section reviews essential components of BBQTs necessary for understanding the research problem we address and the method we introduce. The training process is omitted, as it is not central to this study. All material in this section is drawn from our earlier work [4] and is restated here for completeness.

BBQTs extend QLDTs by incorporating classical Boolean split types into the quantum-inspired framework. The core evaluation principle of QLDTs remains largely intact. Each internal node represents a logical condition formulated in a continuous quantum-inspired logic, where attribute values are interpreted as degrees of truth indicating how strongly a feature expresses a certain trend (e.g., being "high").

BBQTs generalize both QLDTs and classical DTs by allowing three types of splits:

- Q-splits: conditions with gradual truth values
- B-splits: classical DT splits with learned thresholds
- $B_{1/2}$-splits: B-Splits with a fixed threshold at 0.5

This combination preserves several interpretability advantages of QLDTs. Q-splits capture meaningful trends (e.g., "low" hemoglobin or "elevated" glucose in a patient) rather than depending solely on arbitrary numeric cutoffs. As the underlying logic framework limits the usage of attribute per path, BBQTs often remain relatively shallow. In addition, the learning process tends to place the most informative and interpretable conditions near the top of the tree, which supports the interpretation of results.

For the evaluation of a given sample o, each node in a BBQT computes a truth value $[c_{ij}]^o \in [0, 1]$ that quantifies how well the sample satisfies the corresponding condition. A BBQT can be viewed as a collection of paths from the root to each leaf. Within a path, the node values are combined by conjunction, implemented as the product of the node truth values. Class scores are obtained by summing the scores of all paths $\in \mathcal{P}_1$, i. e. paths leading to leaves of the class 1. Then, the score S_1 for class 1 is

$$S_1(o) = [e]^o = \sum_{i \in \mathcal{P}_1} \prod_{j \in \text{path}_i} [c_{ij}]^o.$$

The final prediction is made by comparing S_1 to a learned threshold $\tau \in [0, 1]$.

Unlike classical DTs, which consider only the single path selected by hard splits, BBQTs evaluate *all* paths simultaneously. As mentioned above, this global aggregation can make it more challenging to identify which parts of the tree contributed most to an individual prediction.

IV. Subtree-Based Coloring for Interpretability

This section introduces our subtree-based coloring method, beginning with established decision-tree practices and extending them to BBQ-Trees.

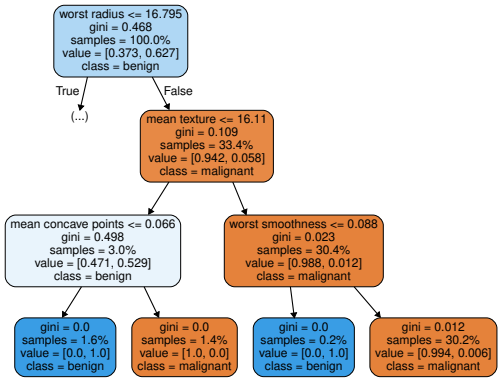


Fig. 1. Right branch of a DT trained on the *Breast Cancer Wisconsin* dataset. Nodes are colored by predicted class and class distribution.

Coloring DTs is a common practice in many machine learning libraries. In fact, standard plotting utilities typically produce color-coded trees by default. For example, scikit-learn's [8] visualization (Figure 1) assigns colors based on the predicted class and the class distribution within each node. In the illustrated example, a DT was trained to predict breast cancer outcomes using the *Breast Cancer Wisconsin* dataset [9]. For compactness, only the right branch of the tree is shown. Orange denotes *malignant* and blue *benign*, while saturation reflects the class distribution in each node. Such visualizations are useful for understanding the global behavior of the tree, but they do not provide a detailed explanation for an individual prediction. Classical DTs, however, operate strictly by traversing a single root-to-leaf path, and this path can be highlighted to indicate which decisions were made for a specific sample. We adapt this idea to BBQTs, where the evaluation is more complex and cannot be represented by a single traversed path.

To visualize BBQTs for a given sample, we compute the score contribution of each subtree to the final class score. These contributions are then mapped to colors using a perceptually uniform colormap (e.g., `viridis`) to ensure that stronger contributions receive visually stronger emphasis.

Intermediate node scores follow the BBQT evaluation scheme. For a node associated with an attribute a, the model evaluates two subtrees: one conjoined with a and one with $\neg a$. The node score is their disjunction,

$$[a]^o * [c_{\text{subtree}_1}]^o + (1 - [a]^o) * [c_{\text{subtree}_2}]^o,$$

where $[\cdot]^o$ denotes the evaluation of a condition for sample o and c_{subtree} represents the condition encoded by a subtree. The same expression contains the edge scores: the first term of the sum corresponds to the edge representing a, and the second to the edge representing $\neg a$. B-split nodes are treated analogously, but using their respective evaluation rule.

Mapping these scores to color intensities produces a visual representation in which users can directly see which parts of the BBQT "push" the prediction higher or lower for an individual patient. Figure 2 presents an exemplary colored BBQT unrelated to any particular dataset. It shows the typical

DOI 10.5162/iCCC2026/P32

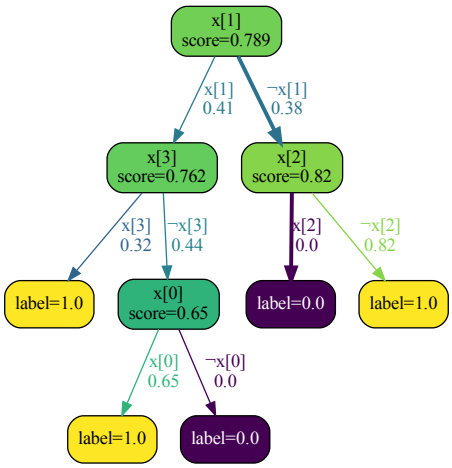


Fig. 2. Example BBQT with heatmap-based coloring scheme

structure and the evaluation for a sample predicted as class 1. Warm colors represent high subtree scores, while cold colors indicate low scores. Here, the root reaches a score of 0.789, reflecting the strong influence of the paths to 1-leaves.

V. Applicability of BBQ-Trees in Medical Settings

In this section, we examine the applicability of BBQTs in medical decision scenarios. We outline the data requirements necessary to use the model effectively and illustrate its behavior on commonly used medical datasets. This motivates improving BBQT interpretability for practitioners who rely on model outputs without being ML experts.

Like other classifiers for tabular data, BBQTs can be applied to structured patient data in settings where binary decisions are required (e.g., disease vs. no disease). Then, each patient corresponds to one data point in a multidimensional feature space. For reliable use of BBQTs, several practical criteria must be fulfilled:

- Balanced datasets for training
- Attribute values scaled to the unit interval $[0, 1]$
- Binary labels
- Non-correlated attributes (Not strict in practice.)

In general, BBQTs perform well when the target class can be represented by distinct regions near the corners of the feature space hypercube.

Table I presents results on four standard medical datasets: *Breast Cancer Wisconsin*, *Pima Indians Diabetes*, *Heart Disease Cleveland*, and *Chronic Kidney Disease* [9]–[12]. While selected metadata is included in the table, additional details are available in the respective descriptions in the cited sources. Across these datasets, accuracy ranges from 79% to 92.5%. For completeness, ROC AUC and other metrics are also reported.

All experiments were conducted using a grid search to identify suitable hyperparameters and evaluated using 4-fold cross validation. The train/test split ratio was set to 0.20. Standard scikit-learn tools were used alongside our custom BBQT implementation.

VI. Demonstration of Subtree-Based Coloring

Figure 3 shows a BBQT trained on the *Autism Screening Data for Toddlers* dataset [13]. This example illustrates clearly how the proposed coloring method supports the interpretation of BBQT predictions. The model is relatively shallow, with depth $d = 4$ plus leaves, yet it already contains $\sum_{i=0}^{d-1} 2^i = 15$ internal nodes. Because a BBQT evaluates all paths globally, even such a small tree might become difficult for non-experts to interpret when presented without visual guidance.

Using the heatmap-based coloring scheme introduced above, nodes and edges with high contribution scores appear in warm colors (yellow), whereas low scores appear in cold colors (purple). In this dataset, most features correspond to binary (yes/no) questions, which results in contributions concentrated at the extremes. This leads to a clear visual separation between influential and non-influential parts of the tree.

The displayed coloring corresponds to the evaluation of a particular sample that the model classified as *ASD positive*. This can be seen immediately at the root node: its color is bright yellow, indicating a high final score (with 1 representing ASD and 0 representing no ASD). Three leaves in the tree predict the positive class, and therefore represent active minterms in the internal CQQL aggregation. By examining which of these leaves contributes most strongly to the root score, we can identify the driving factor behind the decision.

Only the rightmost leaf exhibits a consistently warm-colored path from leaf to root, indicating that this path provided the dominant contribution to the final prediction. The logical condition along this path is:

$$x[5] \wedge x[9] \wedge x[4] \wedge x[10] \geq 0.031.$$

According to the dataset documentation in [13], these features represent:

- $x[5]$: Does the child follow where an adult is looking?
- $x[9]$: Does the child stare at nothing with no apparent purpose?
- $x[4]$: Does the child point?
- $x[10] \geq 0.031$: Normalized age of the child should be greater or equal to 0.031 (~ 2.48 months)

Thus, these features contributed most significantly to the model's positive ASD prediction for this sample. As the dataset's class labels were produced through an automated screening tool called ASDTests [14], they do not possess the level of diagnostic reliability as validated assessments. However, as the purpose of this experiment is to examine the technical applicability of BBQTs rather than to make substantive clinical claims, we hold that this limitation does not compromise our methodological insights.

DOI 10.5162/iCCC2026/P32

TABLE I
EVALUATION RESULTS OF BBQT ON STANDARD MEDICAL DATASETS

Dataset	n_samples	n_features	best_cv_score	accuracy	f1	roc_auc
Breast Cancer Wisconsin	569	30	0.9639	0.9035	0.9197	0.9666
Pima Indians Diabetes	768	8	0.7945	0.7077	0.6666	0.7901
Heart Disease Cleveland	297	13	0.8163	0.7833	0.7346	0.8214
Chronic Kidney Disease	400	24	0.9466	0.925	0.9361	0.94

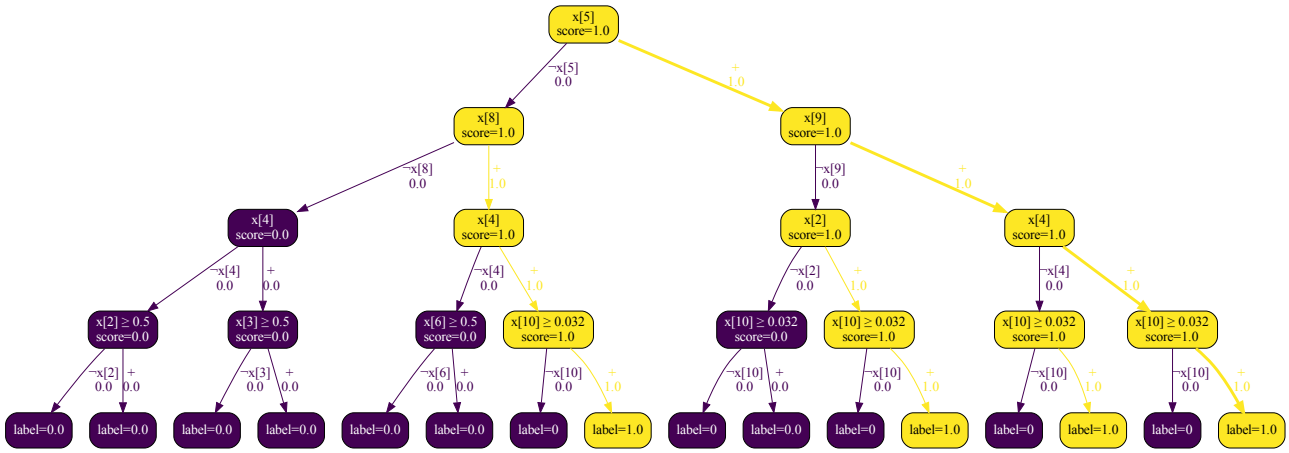


Fig. 3. BBQT for autism diagnosis with 89.63% accuracy, using highlighting for a single prediction

VII. CONCLUSION

This study shows a methodological contribution aimed at improving the interpretability of BBQ-Trees, a class of classification models. While BBQTs can represent complex, non-linear decision boundaries, their evaluation scheme can make it difficult to identify important factors of the decision. We address this by introducing a color-based visualization that maps subtree evaluation scores to color values, for example in a heatmap color scheme. This enables insights into the influences of each decision path in the trained model.

Our experiments on popular medical datasets demonstrate that BBQTs can be applied in settings where binary decisions must be made from structured patient data. Notably, the results are not intended as clinical findings. Instead, they are meant to illustrate the model's applicability and show how the proposed visualization can support interpretability in domains, such as medicine, where understanding decisions is important. The method is general and can be applied to other areas in which transparent model behavior is required.

REFERENCES

[1] A. M. Rahmani, E. Yousefpoor, M. S. Yousefpoor, Z. Mehmood, A. Haider, M. Hosseinzadeh, and R. Ali Naqvi, "Machine learning (ML) in medicine: Review, applications, and challenges," *Mathematics*, vol. 9, no. 22, 2021.

[2] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification and Regression Trees*. New York: CRC Press, 1984.

[3] S. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach, Global Edition*. London: Pearson Education, 2021.

[4] A. Stahl and I. Schmitt, "BBQ-Tree: A unified classifier and regressor combining boolean and quantum logic decisions," *Information Systems*, vol. 136, 2026. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0306437925001188

[5] T. Parrt, D. Johnson, and J. Fairbanks, "dtreeviz: Decision tree visualization library," https://github.com/parrt/dtreeviz, 2019, version accessed: 2025-12-01.

[6] N. Yang and K. Baum, "treeheatr: Heatmaps for decision trees," https://cran.r-project.org/package=treeheatr, 2020, version accessed: 2025-12-01.

[7] C. Molnar, *Interpretable Machine Learning*, 2022, accessed: 2025-11-26.

[8] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[9] M. O. S. N. Wolberg, William and W. Street, "Breast Cancer Wisconsin (Diagnostic)," UCI Machine Learning Repository, 1993, DOI: https://doi.org/10.24432/C5DW2B.

[10] U. M. L. Repository, "Pima indians diabetes database," https://www.kaggle.com/datasets/uciml/pima-indians-diabetes-database, 2016, accessed: 2025-11-29.

[11] S. W. P. M. Janosi, Andras and R. Detrano, "Heart Disease," UCI Machine Learning Repository, 1989, DOI: https://doi.org/10.24432/C52P4X.

[12] S. P. Rubini, L. and P. Eswaran, "Chronic Kidney Disease," UCI Machine Learning Repository, 2015, DOI: https://doi.org/10.24432/C5G020.

[13] F. Abdeljaoued, "Autism screening for toddlers," https://www.kaggle.com/datasets/fabdelja/autism-screening-for-toddlers, 2020, accessed: 2025-11-28.

[14] "Asdtests," https://asdtests.com, accessed: 2025-11-29.

DOI 10.5162/iCCC2026/P33

Securing Custom IEEE 802.11 Based Mesh Networks on ESP32 IoT Modules

Pavlo Mykytyn*, Randolf Rotta*, BTU Cottbus-Senftenberg, Cottbus, Germany
mykytyn@b-tu.de, rottaran@b-tu.de
*The authors contributed equally to this research

Abstract

ESP32-based wireless modules have become a popular and low-cost platform for prototyping IoT and distributed sensing applications. The integrated radios for IEEE 802.11 Wi-Fi, BLE, and IEEE 802.15.4 communication offer a flexible foundation for ad-hoc networking. Existing mesh network implementations for the ESP32 that support encryption for security are limited to an inflexible tree topology of access points. Meanwhile, mesh protocols like B.A.T.M.A.N. for ESP32, which offer fast route repair and throughput optimized route selection, lack implementation of security mechanisms. To close this gap, we first implemented and evaluated a software-based encryption scheme utilizing ECC and AES-GCM for the ESP32 B.A.T.M.A.N mesh. However, the software implementation revealed significant performance limitations, motivating the reverse engineering of the ESP32 radio driver to enable and implement on-the-fly hardware encryption/decryption. This approach makes it possible to secure custom mesh protocols without impairing throughput and processor load.

1 Introduction

Ad-hoc mesh networks provide communication where existing infrastructure is not easily available, does not match the application's communication patterns, or an infrastructure independent solution is desired. Self-organizing mesh protocols can be rapidly deployed and can detect and repair broken routes quickly. Some applications demand mobile deployment and cannot rely on pre-existing communication infrastructure. Examples include temporary field operations [1], autonomous robotic systems [2] or low-power IoT networks in remote areas [3], [4]. Some applications benefit from direct communication across peer devices. For example, cooperative collision avoidance and path planning in mobile swarms relies on periodic position broadcasts and negotiation between nearby vehicles.

The practical usefulness of such self-organizing mesh networks highly depends on their security. Their distributed, decentralized, and sometimes intentionally open nature increases the challenges for ensuring confidentiality, authenticity and integrity of messages as well as detecting and preventing attacks on the link and networking layer, as stated in [5]. Without well-designed security, mesh networks can become a serious safety risk, as described in [6]. Many routing protocols, including the B.A.T.M.A.N (Better Approach To Mobile Ad hoc Networking) [7] mesh protocol, do not incorporate encryption at the network layer, delegating it instead to the application layer. The commonly used Linux kernel implementations can be combined with 802.11s security mechanisms, however, this is not yet available for constrained hardware like the ESP32-S3, C3, or C6 from Espressif Systems. Moreover, Espressif's SDK does support 802.11s, nor does it expose the required mesh peering and key management APIs.

In this work, we extend the implementation of a B.A.T.M.A.N routing protocol with a software-based and hardware-based implementations of AES encryption mechanisms to improve the network security.

2 Background and Related Work

The routing protocol is responsible for discovering and selecting routes between source and destination nodes as well as forwarding packets along these routes. A flooding method is often used for discovering routes. Mesh routing outperforms tree-based routing in high-mobility networks, offering alternative paths for packet transmission from source to destination. The route discovery in these protocols can be on-demand like in the AODV protocol [8], proactive like in the OLSR [9], BABEL [10], B.A.T.M.A.N, or hybrid such as HWMP [11]. Proactive protocols keep routes to every node updated at all times, even if those routes are never used. On-demand protocols only look for a route when it's needed or when one fails.

To the best of our knowledge, there is no open mesh implementation for low-power microcontrollers based on IEEE 802.11 transceivers. To close this gap, we implemented the B.A.T.M.A.N protocol and the required software stack on the Espressif ESP32 platform [4].

2.1 Attack Vectors on B.A.T.M.A.N Meshes

B.A.T.M.A.N-adv extends the original B.A.T.M.A.N protocol by shifting routing from the network layer to the data link layer and forwards frames based on MAC addresses instead of IP addresses. This design simplifies multi-hop communication, making it better suited for lightweight, dynamic mesh deployments.

However, even the advanced version of B.A.T.M.A.N does not include any built-in encryption, authentication, or integrity checks. All control messages, such as originator messages and topology updates are sent in plaintext. An eavesdropper can, therefore, learn every node's address and link metrics, effectively revealing the mesh topology. Because there is no cryptographic validation, any node can inject or spoof routing packets. In practice this exposes the mesh to:

Eavesdropping attacks, where a passive listener can learn the entire network topology and eavesdrop on the exchanged messages.
Flooding attacks, where malicious nodes can flood the network with fake originator messages and malformed frames to overwhelm the legitimate routing updates.
Blackhole attacks, where an attacker falsely announces the best route to a destination, claiming to be the next hop, and drawing traffic into a black hole.
Wormhole attacks, where legitimate packets are relayed over large distances in order to create a false sense of proximity and attract more traffic for eavesdropping.

2.2 Wi-Fi WPA3 and 802.11s Security

WPA3-Personal uses Simultaneous Authentication of Equals (SAE), a mechanism that was originally developed for 802.11s wireless mesh networks [13]. Based on SAE, a pairwise secured channel is established and then used to establish a Group Master Key (GMK) and Pairwise Master Key (PMK). A Group Temporal Key (GTK) is derived from the GMK and is used to encrypt broadcast frames. A Pairwise Transient Key (PTK) is derived from the PMK and parts of it are used to derive the Temporal Key (TK) that is used to encrypt unicast frames. On Linux-based single-board computers, 802.11s based peer authentication and encryption can be combined with custom mesh protocols such as BATMAN.
Although the ESP32 firmware does not implement 802.11s routing, the WPA3 personal and WPA3 security is implemented and the hardware supports the respective symmetric ciphers. The encryption capabilities of the radio hardware were first described by [14] based on reverse engineering of the Wi-Fi SoftMAC driver. Espressif offers a proprietary, closed-source mesh called ESP-MESH, based on the legacy WPA2 standard and tree-based routing [15]. Mesh security also depends on managing group membership. WPA3-Personal lacks any standard method to revoke keys or remove peers. Once connected, a node retains the key indefinitely. This is a major issue for any mesh with shifting membership. WPA3-Enterprise addresses this through a central authentication service based on RADIUS. However, forwarding the distributed association requests via a resource-constrained mesh network to this central authenticator is susceptible to denial-of-service attacks.

3 Authenticated Key Agreement and Software-based Encryption

To secure the communication in a mesh network, we first implemented and tested a mechanism that uses Elliptic Curve Cryptography (ECC) combined with a software implementation of symmetric encryption using Advanced Encryption Standard (AES) in Galois-Counter Mode (GCM). We have already partially discussed this approach in our earlier works [3], [16]. The key agreement process is conducted pairwise to establish a session key. We use Elliptic Curve Diffie-Hellman (ECDH) key exchange between each node and the base station. However, ECDH alone does not guarantee authenticity and does not protect against Man-in-the-Middle attacks. Therefore, to authenticate the key agreement process, both the base station and each node are provided with their own Elliptic Curve Digital Signature Algorithm (ECDSA) keys, stored in the non-volatile storage (NVS). Once the pairwise ECDH key agreement is initiated, the base station signs its public ECDH key with its private ECDSA key and distributes it to each node in the mesh. Upon receipt, each party verifies the authenticity of the ECDH public keys and then uses them to compute a shared secret and generate a session key.
Once a session key has been established, it is used to generate further pairwise keys using a Key Derivation Function (KDF) and symmetrically encrypt the communication and ensure forward secrecy. The multi-hop mesh communication is encrypted using software-implementation of the 256-bit AES-GCM, based on the open-source Mbed-TLS library [17] with hardware-specific optimizations for the ESP32. It provides authenticity and integrity by performing Galois field multiplication of the ciphertext with a hash of the plaintext and associated data. The throughput and performance testing of this implementation are presented in the next section.

3.1 Throughput Benchmarks

We used ESP32-S3 modules with 16 MB of flash memory, 8 MB of external SPI-RAM, and an on-board PCB antenna. They were programmed using ESP-IDF v5.1.4 with the CPU frequency set to 240 MHz, flash memory frequency to 120 MHz, and SPI-RAM frequency also to 120 MHz, ensuring optimal performance for our experiments.
Seven nodes were arranged along a hallway, spaced 10 meters apart from each other, to create a mesh network as presented on the Figure 1. To enforce multi-hop communication, the transmit power was reduced to 8 dBm. The longest communication route spanned 4 hops. The throughput was measured by continuously sending 1400-byte blocks for 5 seconds, with the number of received bytes recorded every 250 ms on the receiving node, resulting in 20 throughput samples for each connection type.

The first configuration ran both encryption/decryption with AES-GCM and a 256-bit key and routing tasks on core 0. It is marked with blue color in the Figure 2. In the second setup, encryption/decryption was done on core 0, while the routing task ran on core 1. It is marked with orange color in the Figure 2. As a baseline, the third setup used no encryption and is marked with green color on the Figure 2.

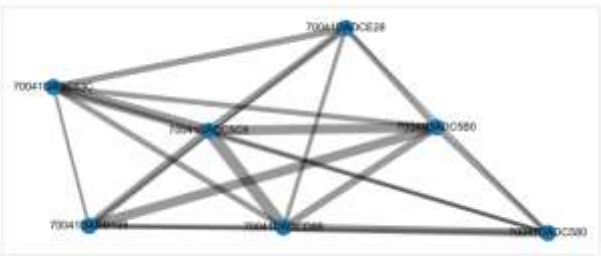


Figure 1 Node layout configuration of the mesh network

DOI 10.5162/iCCC2026/P33

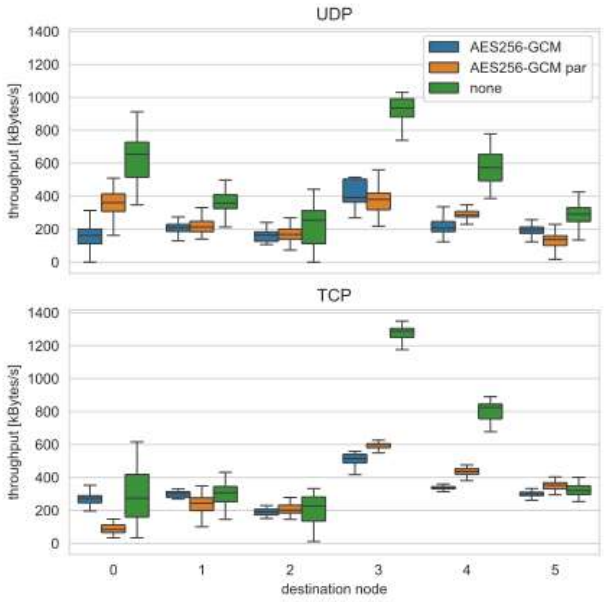


Figure 2 Averaged throughput of up to 5 hops, over UDP and TCP connections.

The software-based encryption shows a significant reduction of throughput on high-throughput single-hop routes: Without encryption, the highest measured throughput between a node pair was around 1300 kBytes/s and with encryption, the throughput between the same pair dropped to around 400 kBytes/s. On longer multi-hop routes, the measured throughput did not exceed 400 kBytes/s without encryption. Fortunately, the throughput with encryption did not drop significantly. This can be explained due to the additional processing overhead caused by encryption and decryption mostly overlapping with delays caused by the collisions between incoming and forwarded frames.

4 Hardware-Based Encryption in ESP32 Wi-Fi Radio

Newer versions of the ESP32 SoC, such as S3 and C6 support WPA3-Enterprise and, hence, hardware-based encryption for AES-GCMP with 256-bit keys. However, because of the closed source radio driver, this was not accessible to custom mesh network implementations. This section summarizes insights from partial reverse-engineering and proposes a programming interface to simplify the integration into custom network stacks.

4.1 Encryption Key Slot Architecture

The 802.11 radio integrated in the ESP32 SoC can perform encryption and decryption of frames on the fly. For this, the symmetric temporal key needs to be placed into one of the 25 available hardware key slots. Each key slot also contains the configuration of the cipher algorithm, a MAC address, the 2-bit WPA key ID to select the right key for incoming frames, the hardware interface (STA/AP) the key applies to, and a few configuration flags to match, for example, broadcast frames. The WPA key ID is used for periodic key rollover.

A few internal undocumented functions can be used to install own keys. However, they assume, that group temporal keys for broadcast frames are always placed in the first four slots. Higher layers of the driver manage the slot assignment and periodic key rollover. The station interface needs just four slots: two for the group key and two for the pairwise key towards the access point. The access point interface needs two slots for the group temporal key and two slots for each associated station. Thus, at most 9 stations can connect to the access point.

The frame format follows the IEEE 802.11 standard. If the Protected Frame flag is set in the header, an 8-byte WPA header follows, which contains the unique packet number and WPA key ID. Then follows the payload and, finally, a 16 bytes Message Integrity Check (MIC) code. The protection flag, WPA header, and MIC area have to be inserted manually before transmission. The radio will then compute and verify the MIC transparently.

The radio manages multiple pending transmission and the hardware key slot number needs to be configured in the respective transmit slot in order to apply the desired encryption key. The low-level transmit functions read this key slot number from the frame's metadata. The documented transmit function for custom frames does not configure this part of the metadata, but it is also not changed later.

Decryption of received frames is done transparently. For unicast frames, the transmitter address from the frame header and the WPA key ID are used to find a matching entry in the hardware key slots. For broadcast frames, the first key slot that matches the interface (STA/AP) and WPA key ID is used while the frame's BSSID address is ignored. When the cryptographic verification fails, the frame is still acknowledged to the transmitter but discarded. If no hardware key slot matches, the frame is not decrypted but still reported via the monitor mode callback.

4.2 Proposed Programming Interface for Custom Mesh Networks

The programming interface should make it possible to manage a few reserved key slots for associated peers and broadcasts, select the right key for outgoing frames, and identify received frames that were not decrypted because no key slot matched. A low-level approach would be to provide direct access to the reserved key slots in order to directly install and configure keys. In this case, the transmit function for custom frames needs an additional argument to select the encryption key slot. We followed this approach for our own experiments.

A more comfortable interface would allow to manage a list of associated peers based on their MAC address. Then, the assignment to key slots would be handled internally and the transmit function transparently looks up the correct key slot for each outgoing frame. An example for such interface is the peer management in Espressif's ESP-NOW protocol, which allows to configure a key for each associated peer.

Although more comfortable, this interface has a couple of drawbacks. Most importantly, different strategies exist for the key rollover and respective use of multiple key slots. A

high-level management would mandate a specific strategy. For example, the mesh protocol can use clock synchronization to coordinate key rollovers such that a single slot per peer is sufficient. Secondly, the mesh routing layer selects the next-hop peer for forwarding and, hence, already knows the key slot such that the internal search would not be necessary.

Finally, the 802.11s standard uses a different Group Temporal Key on each mesh node for broadcast frames. Thus, it would be necessary to also manage this key for each peer. However, according to our reverse engineering efforts, the hardware does not support multiple broadcast keys because of the key slot selection process for received broadcast frames. In addition, even with a workaround, this needs a second installed broadcast key for each peer and, thus, will reduce the maximum number of peers from 20 to 10, which would severely restrict the mesh protocol. Instead, we propose to use a mesh-wide group temporal key for broadcasts in combination with time synchronized key rollover.

5 Conclusions

ESP32 modules have become a popular low-cost platform for IoT applications. The introduction of 802.11-based self-organizing mesh networks enables new applications. These applications require confidentiality, integrity, and availability in the wireless communication. This paper shows that software and hardware encryption can be integrated into custom mesh implementations. However, software-based encryption causes additional processing overhead that significantly reduces the peak throughput. This can be overcome by making use of the radio's built-in hardware cryptography.

References

[1] A. G. Q. Aquino, A. H. Ballado, and A. V. Bautista, "Implementing a Wireless Sensor Network with Multiple Arduino-Based Farming Multi-Sensor Tool to Monitor a Small Farm Area Using ESP32 Microcontroller Board," in *2021 IEEE 13th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management (HNICEM)*, Nov. 2021, pp. 1–6. doi: 10.1109/HNICEM54116.2021.9731989.

[2] "Heterogeneous Ground-Air Autonomous Vehicle Networking in Austere Environments: Practical Implementation of a Mesh Network in the DARPA Subterranean Challenge." Accessed: Nov. 10, 2025. [Online]. Available: https://www.computer.org/csdl/proceedings-article/dcoss/2022/951200a261/1GBSOhnaUvu

[3] R. Rotta and P. Mykytyn, *Secure Multi-hop Telemetry Broadcasts for UAV Swarm Communication*. 2024. doi: 10.26127/BTUOpen-6637.

[4] R. Rotta, J. Schulz, B. Naumann, N. S. Chatharajupalli, J. Nolte, and M. Werner, "B.A.T.M.A.N. Mesh Networking on ESP32's 802.11," in *2024 IEEE 49th Conference on Local Computer Networks (LCN)*, Oct. 2024, pp. 1–7. doi: 10.1109/LCN60385.2024.10639698.

[5] M. A. Lopez, M. Baddeley, W. T. Lunardi, A. Pandey, and J.-P. Giacalone, "Towards Secure Wireless Mesh Networks for UAV Swarm Connectivity: Current Threats, Research, and Opportunities," in *2021 17th International Conference on Distributed Computing in Sensor Systems (DCOSS)*, July 2021, pp. 319–326. doi: 10.1109/DCOSS52077.2021.00059.

[6] M. R. Albrecht, J. Blasco, R. B. Jensen, and L. Mareková, "Mesh Messaging in Large-scale Protests: Breaking Bridgefy," 2021, 2021/214. Accessed: Nov. 10, 2025. [Online]. Available: https://eprint.iacr.org/2021/214

[7] "Doc-overview - batman-adv - Open Mesh." Accessed: Nov. 10, 2025. [Online]. Available: https://www.open-mesh.org/projects/batman-adv/wiki

[8] S. R. Das, C. E. Perkins, and E. M. Belding-Royer, "Ad hoc On-Demand Distance Vector (AODV) Routing," Internet Engineering Task Force, Request for Comments RFC 3561, July 2003. doi: 10.17487/RFC3561.

[9] T. H. Clausen and P. Jacquet, "Optimized Link State Routing Protocol (OLSR)," Internet Engineering Task Force, Request for Comments RFC 3626, Oct. 2003. doi: 10.17487/RFC3626.

[10] J. Chroboczek and D. Schinazi, "The Babel Routing Protocol," Internet Engineering Task Force, Request for Comments RFC 8966, Jan. 2021. doi: 10.17487/RFC8966.

[11] M. A. Ng and K.-L. A. Yau, "An energy-efficient Hybrid Wireless Mesh Protocol (HWMP) for IEEE 802.11s mesh networks," in *2013 IEEE International Conference on Control System, Computing and Engineering*, Nov. 2013, pp. 17–21. doi: 10.1109/ICCSCE.2013.6719925.

[12] "batman-adv — The Linux Kernel documentation." Accessed: Nov. 11, 2025. [Online]. Available: https://www.kernel.org/doc/html/v4.15/networking/batman-adv.html

[13] "IEEE 802.11s — Linux Wireless documentation." Accessed: Nov. 14, 2025. [Online]. Available: https://wireless.docs.kernel.org/en/latest/en/developers/documentation/ieee80211/802.11s.html

[14] J. Devreker, "Reverse engineering ESP32 Wi-Fi driver: the road ahead," ESP32 open MAC. Accessed: Nov. 14, 2025. [Online]. Available: https://esp32-open-mac.be/posts/0005-the-road-ahead/

[15] "ESP-WIFI-MESH - ESP32 - — ESP-IDF Programming Guide v5.5.1 documentation." Accessed: Nov. 14, 2025. [Online]. Available: https://docs.espressif.com/projects/esp-idf/en/stable/esp32/api-guides/esp-wifi-mesh.html

[16] P. Mykytyn, M. Brzozowski, Z. Dyka, and P. Langendoerfer, *Towards Secure and Reliable Heterogeneous Real-time Telemetry Communication in Autonomous UAV Swarms*. 2024, p. 165. doi: 10.5162/iCCC2024/P15.

[17] "Mbed TLS · GitHub." Accessed: Nov. 12, 2025. [Online]. Available: https://github.com/Mbed-TLS

DOI 10.5162/iCCC2026/P34

A 5G power-hardware-in-the-loop testbed for smart low-inertia power systems applications

Rolando Cortés Martínez[1], **Nandavardhan Reddy Ramireddy**[2], **Eric William Zurita**[1], **and Johannes Schiffer**[1,3]

[1]Control Systems and Network Control Technology Group, Brandenburg University of Technology Cottbus-Senftenberg (BTU C-S), 03046 Cottbus, Germany
[2]Institute for Computer Science, Brandenburg University of Technology Cottbus-Senftenberg (BTU C-S), 03046 Cottbus, Germany
[3]Fraunhofer IEG, Fraunhofer Research Institution for Energy Infrastructures and Geotechnologies, 03046 Cottbus, Germany

Correspondence: Rolando Cortés Martínez (cortesma@b-tu.de)

Abstract. The use of 5G communication technologies is broadly expanding in different areas. Power systems are a highly promising area of application, where the advantages of 5G can play a significant role. Some of the 5G outstanding features are good reliability, low latency, and low jitter, among others. However, there is a challenge in incorporating new technologies into highly complex and extremely crucial infrastructures. In this work, we present a testbed that tackles this problem by combining a power-hardware-in-the-loop environment with a 5G communication network. We present preliminary results that validate the approach by showing relevant performance indexes around 5G.

1 Introduction

The increased addition of renewable energy sources (RES) into the electrical grids has yielded the need for experimental facilities that allow exploring the new control paradigms involved in such low-inertia power systems (see e.g., Krishna et al. (2022)). For such purposes, power-hardware-in-the-loop (PHIL) testbeds offer good flexibility and safety, making them ideal for testing a variety of scenarios found in the industry before real implementation. Thanks to the use of PHIL facilities, many scenarios can be reproduced under controlled conditions. A natural capability is to emulate the mechanical and electrical dynamics of RES (see e.g., Fadaeinedjad et al. (2008), Chiniforoosh et al. (2010)).

Typical communication solutions in real power systems consist of using large wired cables based on industrial Ethernet or costly optical fiber. However, Porcu et al. (2021) show that to improve flexibility and reliability, the industry is exploring the use of 5G technology, which also offers a good range of coverage and massive device capacity. Xu et al. (2020) emphasize the characteristics that make 5G ideal for these applications, such as reliability, low latency, massive capacity in terms of the number of nodes, high transmission speeds, and edge computing, among others. Porcu et al. (2022) propose a scheme to emulate distributed generation control with 5G communication. For this, a grid emulator (GE) is connected to other power sources to contribute to the balance of frequency and power supply. The sensor and actuation signals from one RES can be flexibly shared with the operator control station.

The current PHIL laboratories lack 5G technology (e.g. von Jouanne et al. (2023)). However, this is crucial to incorporate it since the stability of the testbed is affected by the presence of time delays and jitter in the communication link. With this need in focus, we modify the PHIL testbed developed at the Control Systems and Network Control Technology department at the Brandenburg University of Technology (BTU), Cottbus-Senftenberg, Germany (see Krishna et al. (2022)). The purpose of this setup is to establish a controlled experimental environment for evaluating 5G communication in power network applications, where the PHIL platform emulates the electrical grid and assess the system's behavior under 5G operation. We present the following contributions: 1) We design the architecture to set up the testbed for incorporating 5G devices into the PHIL laboratory, 2) we measure key performance indicators (KPIs) of every 5G equipment to show its status, and 3) we show the capabilities of the testbed

DOI 10.5162/iCCC2026/P34

by running the 5G and PHIL combined for future applications.

2 Design of experimental setup

2.1 PHIL testbed

The PHIL facilities used in this work consist mainly of four distributed generation units (DGUs), one grid emulator (GE) unit, a power line emulation cabinet, a control supervisory computer and the input grid of the University (see e.g., Krishna et al. (2022)). The system has a total power of 110 kW and can convert the 230/400 AC 50 Hz three-phase input power to variable AC or DC voltages with output frequencies from 0 to 400 Hz. Each DGU is controlled by a real-time computer (RTC), which runs embedded code generated under the Matlab/Simulink environment. Thus, each RTC can be used to emulate different electrical dynamics (e.g., storage elements, RESs, virtual synchronous generators, controllers, etc).

2.2 5G and PHIL integration

In this work, we use the 5G HAT SIM8200EA-M2 kit by Waveshare. We refer to this single equipment as a 5G user equipment (5G-UE). This is integrated by a Raspberry Pi 4 (RPi) as an embedded computer, a Raspberry HAT that incorporates the 5G modem SIM8200EA and five antennas. It supports 5G/4G/3G and stand-alone (SA) and non-stand-alone (NSA) networks. The kit features Ethernet and USB ports, as well as general-purpose input/output (GPIO) ports. The combined capabilities of PHIL hardware and the 5G hardware (5G-PHIL) is shown in Fig. 1. The 5G-UEs are connected through a dedicated local Ethernet network to the PHIL hardware. An Ethernet interface is combined with the user datagram protocol (UDP) as a transport layer. The 5G connectivity is provided by a private stand-alone (SA) 5G network in the frequency band n78, with the server antennas located inside the PHIL laboratory.

The use of Ethernet as a local network in addition to the 5G network does not compromise any of the advantages of the 5G network, as Ethernet, combined with the UDP protocol, offers lower delays and faster transmission speeds than 5G radio signals. The proposed interconnection aligns with real-world scenarios, where power systemd hardware can commonly communicate using standard protocols, such as IEC 61850, PMU IEEE C37.118, DNP3, SCADA DNP3, Modbus, CAN (see e.g. von Jouanne et al. (2023)). The main difference, and for the convenience of the laboratory, is that we provide a single local network for all the power devices. The topology in the local network, however, is configured such that only one 5G-UE is paired with one corresponding DGU or GE hardware. The pairing is achieved by identifying and matching the IP addresses during transmission.

On the radio signal side, every 5G-UE is configured to have a fixed IP address for sending and receiving data with other 5G-UEs. In Fig. 1, each 5G-UE has two fixed IP addresses, one for the Ethernet interface and the other for the 5G interface.

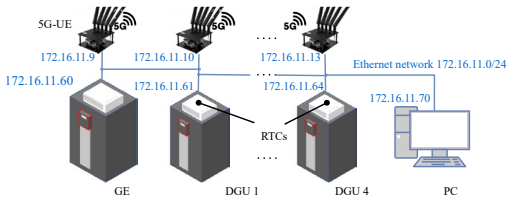


Figure 1. General diagram of the 5G-PHIL setup.

Application communication protocol. Ethernet and 5G constitute the physical communication layers while UDP is the transportation layer, however for the definition of the user data packages we use the message queuing telemetry transport (MQTT) protocol. MQTT is easy to implement and offers flexibility in the definition of the network topologies. In this way, it is possible to turn on and off each one of the links in the network, which is important to test stability conditions in the network. It operates under a *client*, *topic*, *subscriber*, and *publisher* scheme, where the number of clients in the network can be easily scaled up.

2.3 Mode of operation

The testbed has the capability of testing the communication part with or without utilizing the PHIL hardware. This feature facilitates debug processing. During the debug process, instead of connecting all the RTCs to the 5G-UEs, the 5G-UEs can be connected to a single computer in the Ethernet network. This operation, however, lacks real-time capabilities. This computer also serves for editing and development of the codes inside the 5G-UEs. During this process, the computer utilizes the secure shell (SSH) protocol over Ethernet to access the Linux system remotely from other systems, such as Windows.

When the 5G-UEs are paired with the GE and the DGUs, the main function of the PC is to edit, upload, and execute code dedicated to each RTC and its corresponding power system device (see, e.g., Krishna et al. (2022)).

The MQTT client publishes its own set of *topics* related to the physical measurements of the DGU (e.g., voltages, currents, frequency, control variables, etc.), the names of the *topics* follow the form *cliXX_Y*, where *XX* is the number of the MQTT client and *Y* is the number of the *topic* related to that client. For the MQTT broker, which can also act as a client, the *topic* name is *broker_Y*. The name of the *topic* is not transmitted through the UDP-Ethernet to maintain simplicity in this stage. For these reasons, a predefined number of *topics* and a relationship between signals and *topics* are

DOI 10.5162/iCCC2026/P34

necessary. An example of the system sending a signal from the RTC of the GE to the RTC of DGU 1 is shown in Fig. 2. The data, which could be an electrical measurement of the emulated Grid, is sent by Ethernet UDP to the RPi of the 5G-UE 1. Then, the data is transformed into MQTT *topics* and is published using the 5G network. Any RPi subscribed to the published *topics* (RPi 2 in Fig. 2) receives the data and packs it to send it again through the Ethernet UDP network to the target RTC.

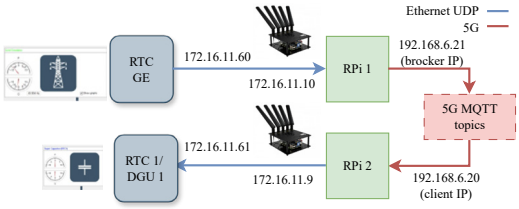


Figure 2. Example for sending information from one RTC to another through 5G-UEs.

Software layer. The five RPis work with the Ubuntu 20.04 server version, with the robotic operating system (ROS) 2 foxy as middleware, the Paho MQTT libraries, and the SIM8200 driver. Our main application code is made in Python language under a ROS 2 node. To understand the internal process of the 5G-UEs in example shown in 2, consider the diagram in Fig. 3. The RPi 1 receives the UDP stream and splits it into individual floating-point values. Then it publishes them into *topics* using multiple threads. This parallelizes the process, where every publication implies a transmission time delay. The RPi-2 subscribes to the *topics*, packages, and streams them over Ethernet using UDP. The pro-

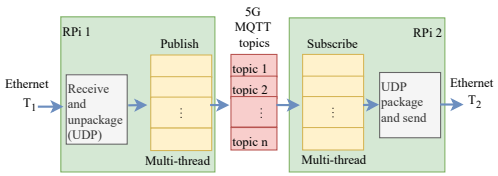


Figure 3. MQTT *topics* publication-subscription process.

cesses are event-handled (as soon as a new UDP package is detected, concurrent MQTT *topics* publication is done). Analogously, as soon as new data arrives in a *topic*, the subscription to it ensures immediate reception of the information to proceed to send it through UDP. On the RTC's side, the process of sending and receiving UDP packages is repeated at regular periods $T_1 = 10$ ms and T_2 = 1 ms, respectively. These values are selected to avoid buffer overflow.

The full or partial set of MQTT *topics* is available to every MQTT client. The client only needs to subscribe to the target *topics*. Thus, the active subscriptions define the topology of the communication. In this way, the 5G-PHIL operator can choose to simulate any kind of topology, such as

Table 1. Ethernet RTCs network configuration (IP address[1]):(port).

Flow	RTC 1	RTC 2	RTC 3	RTC 4
Send	10:7000	11:7000	12:7000	13:7000
Receive	10:8001	10:8002	10:8003	10:8004

[1] Network 172.16.11.0

directed, undirected, connected, strongly connected, etc. To automate the subscriptions based on the desired topology, an initial *topic* publication containing the topology information can be shared through the whole network.

3 System performance analysis

The purpose of this analysis is to assess the 5G performance and to show its suitability for power grid applications. Table 1 shows the configuration of the IP addresses and ports for the RTCs. Every 5G-UE is bound to every RTC by using the same respective pair (IP, port). For convenience, every receiving port is different from each other to maintain compatibility during the communication debug process (Matlab/Simulink supports only one port bound to a single IP address).

The list of *topics* comprises 20 elements with names from *broker_1* to *broker_4* and *cli20_1* to *cli24_4*. Every client is subscribed to all the *topics* in the list except for its own *topics*. Every *topic* corresponds to a Matlab/Simulink variable of type *single* (floating-point value, or four bytes).

To evaluate the performance of the 5G-UEs, we use radio signal measurements such as reference signal received power (RSRP), reference signal received quality (RSRQ) and signal-to-noise ratio (SNR) as well as network measurements such as the round trip time (RTT) between the edge 5G server and the 5G-UE, and the jitter (variation in the RTT). Additionally, by using application-layer data, we measure the latency in the transmissions under the normal PHIL operation mode. The RTT and jitter are obtained during a test carried out during one hour for the five 5G-UEs, a one-minute sample for one device is presented in Fig. 4, and Table 2 summarizes the average values obtained for all devices during the test. As can be seen, the values are consistent with those reported in the literature for indoor environments (see e.g., Singh et al. (2024)) Note that an RTT between UEs is twice the RTT w.r.t. the 5G server.

On a second test, we measure the downlink and uplink throughputs under different communication tasks (see Fig. 5). During the intervals t_1=(17:38, 17:46) and t_2=(17:46, 17:50), ping tasks with periods of 0.1 s and 0.01 s were executed, respectively. In a third test, we measure the 5G-PHIL latency between two RTCs (e.g., Fig. 2). To simplify the task, we utilize a single RTC. A signal $s_t(t) = \text{ramp}(t)$ is transmitted from RTC 1 and received as a delayed sig-

DOI 10.5162/iCCC2026/P34

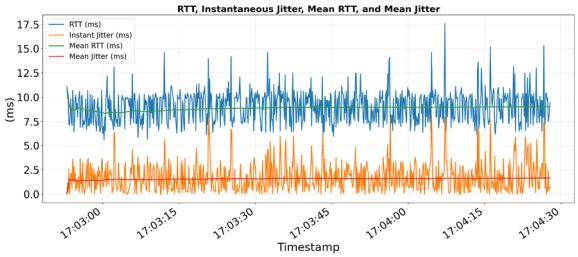


Figure 4. RTT and jitter for the MQTT broker.

Table 2. Network and radio signal KPIs (dB) in the five 5G-UEs.

KPI	broker	client 20[1]	client 22[1]	client 23[1]	client 24[1]
RTT	9.01	9.17	9.25	8.57	9.46
Jitter	1.66	1.56	1.58	1.53	1.71
RSRP	-64	-65	-70	-65	-66
RSRQ	-11	-11	-11	-11	-11
SNR	40	40	40	33.5	36.5

[1] The numbers correspond to the MQTT client identification, same as its 5G IP address and same as *topic* names.

nal $s_r(t) = s_t(t-\tau(t))$, where $\tau(t)$ is the accumulated time-varying delay from: publication from RPi 1, subscription in RPi 2, back publication in RPi 2, subscription in RPi 1, and UDP transmission to the original RTC. The latency measured under this scheme is shown in Fig. 6, with a mean value of $\bar{\tau}(t)$ =91.3 ms. The mean of the effective delay between RTCs is then $\bar{\tau}(t)/2$ =45.7 ms. Such a delay is still acceptable for several real-time power system applications compared to traditional wired solutions, where the delays can go from 50 ms to 130 ms (e.g., Zacharia et al. (2020)).

4 Conclusions

The integration of 5G technology with the PHIL facilities into the 5G-PHIL is a promising approach for representing power system scenarios in a more realistic environment where 5G communication can be effectively implemented. The KPI measurements obtained are consistent with the existing literature. Moreover, radio signal measurement levels such as the signal power, signal interference, and noise ratio are within an excellent range of operations.

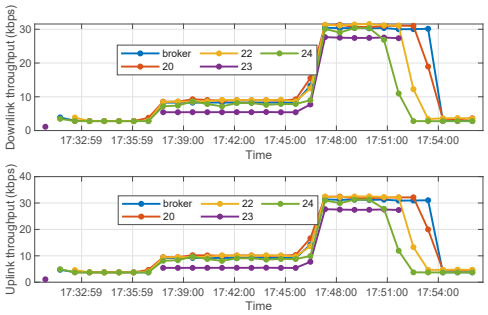


Figure 5. Throughput for uplink and downlink during second test for the five 5G-UEs (MQTT clients 20, 22, 23, and 24).

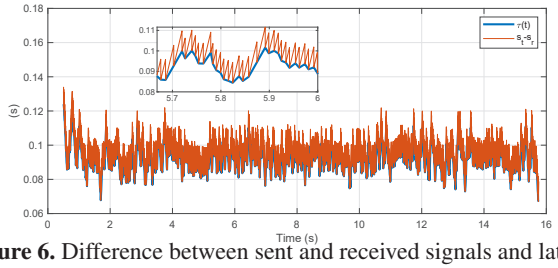


Figure 6. Difference between sent and received signals and latency

References

Chiniforoosh, S., Jatskevich, J., Yazdani, A., Sood, V., Dinavahi, V., Martinez, J. A., and Ramirez, A.: Definitions and applications of dynamic average models for analysis of power systems, IEEE Transactions on Power Delivery, 25, 2655–2669, https://doi.org/10.1109/TPWRD.2010.2043859, 2010.

Fadaeinedjad, R., Moallem, M., and Moschopoulos, G.: Simulation of a wind turbine with doubly fed induction generator by FAST and simulink, IEEE Transactions on Energy Conversion, 23, 690–700, https://doi.org/10.1109/TEC.2007.914307, 2008.

Krishna, A., Jaramillo-Cajica, I., Auer, S., and Schiffer, J.: A power-hardware-in-the-loop testbed for intelligent operation and control of low-inertia power systems, at-Automatisierungstechnik, 70, 1084–1095, 2022.

Porcu, D., Chochliouros, I. P., Castro, S., Fiorentino, G., Costa, R., Nodaros, D., Koumaras, V., Brasca, F., Di Pietro, N., Papaioannou, G., et al.: 5G communications as “enabler” for smart power grids: the case of the Smart5Grid project, in: IFIP International Conference on Artificial Intelligence Applications and Innovations, pp. 7–20, Springer, 2021.

Porcu, D., Castro, S., Otura, B., Encinar, P., Chochliouros, I., Ciornei, I., Hadjidemetriou, L., Ellinas, G., Santiago, R., Grigoriou, E., et al.: Demonstration of 5G solutions for smart energy grids of the future: a perspective of the Smart5Grid project, Energies, 15, 839, 2022.

Singh, B. C., Diaz, R., and Shetty, S.: Evaluating the Performance of 5G NR in Indoor Environments: An Experimental Study, Procedia Computer Science, 241, 16–23, 2024.

von Jouanne, A., Agamloh, E., and Yokochi, A.: Power hardware-in-the-loop (PHIL): A review to advance smart inverter-based grid-edge solutions, Energies, 16, 916, 2023.

Xu, D., Zhou, A., Zhang, X., Wang, G., Liu, X., An, C., Shi, Y., Liu, L., and Ma, H.: Understanding operational 5G: A first measurement study on its coverage, performance and energy consumption, in: Proceedings of the Annual Conference of the ACM, pp. 479–494, 2020.

Zacharia, L., Asprou, M., and Kyriakides, E.: Measurement Errors and Delays on Wide-Area Control Based on IEEE Std C37.118.1-2011: Impact and Compensation, IEEE Systems Journal, 14, 422–432, https://doi.org/10.1109/JSYST.2019.2918841, 2020.

DOI 10.5162/iCCC2026/P35

Step by Step from Modbus to secure IoT

Reinhardt Karnapke
Perinet GmbH
Siemens-Halske-Ring 2
03046 Cottbus
reinhardt.karnapke@perinet.io

Karsten Walther
Perinet GmbH
Rudower Chaussee 29
12489 Berlin
karsten.walther@perinet.io

Abstract

Many industrial deployments have been realized using Modbus technology in the last 46 years. However, new challenges like the Data Act of the European Union, security issues in deployments as well as a general need for more Data availability up in the Enterprise System lead to demands that traditional Modbus deployments simply cannot fulfill.
Fulfilling these new requirements seems to mandate a complete rebuild of legacy deployments. However, no one wants to stop production and incur large downtimes for entire sections of factories at once.
In this paper we introduce a step-by-step approach for switching from classic Modbus or Modbus TCP to a secure IoT based operation with minimal to no downtime.

1 Introduction

Despite having been designed almost 50 years ago with PLCs in mind, Modbus1 remains very relevant today. Many factory installations are built using field busses, with Modbus being most prominent among them.
However, even though the installations still work, there are some additional challenges that have arisen over the years, which cannot be addressed by legacy Modbus installations very well. These include, among others, regulatory issues, accessibility from the enterprise system and security concerns.

Having the ability to access data from the enterprise system would in itself be rewarding. Whether the rewards are high enough to warrant the costs depended on the factory owner, until recently.
However, with the Data Act of the European Union[1] that came into effect on January 11, 2024, this changed.
Since September 12th, 2025, this Data Act grants all customers the right to demand access to their product related Data. By September 12th, 2026 all new products placed on the European market will have the have included Data accessibility already during the draft phase.

Security has long been realized with physical access to factories in mind, not with access to Data. Therefore, classic Modbus deployments transmit Data unencrypted. However, nowadays a lot of operations are outsourced, meaning that engineers from different companies, in some deployments even customers, can get close to the machines and the Modbus cables, usually RS-485. As this data is transmitted unencrypted, anyone with physical access can just attach a Data sniffer and collect Data.
There is the theoretical option of using TLS over the existing Modbus. However, this would decrease the Bandwidth available to the already running applications. Even if this could be tolerated, the even bigger drawback is that the Firmware of all participating machines would have to be updated, and simultaneously.

Even though Data availability and security are big concerns, it is simply not practicable to remove whole installations and replace them with other, secure approaches as this would result in huge downtimes of the deployments. Therefore, a way to gradually introduce Data collection and security into existing Modbus deployments without interrupting day to day operation, or at least with minimal interference, is needed.

In this paper we describe our approach for retrofitting existing industrial Modbus deployments with security and connections to the Enterprise System without shutting down the operation of the whole deployment. Instead, we focus on a step by step approach, adding additional communication capabilities and replacing elements only when they fail or are not in use, leading to undisrupted operation.

2 Step by Step from Modbus to secure IoT

Figure 1 shows a typical Modbus deployment as can be found in any number of factories today.

Local controllers are connected to Modbus TCP/RTU gateways, which are in turn connected with RS-485 cables on which the Modbus devices reside. Currently, there is no connection to the Enterprise System.

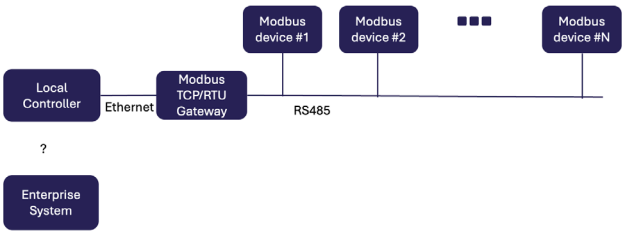


Figure 1 Current Situation in many Industrial Deployments

DOI 10.5162/iCCC2026/P35

2.1 Minimal approach: Insert Secure Data Sniffer

For the minimal approach we focus on gathering Data from the Modbus and making it available to the Enterprise System and, therefore, the customer, when requested.
For this, we can make use of the currently still insecure Data transmission on the Modbus by inserting a Modbus Sniffer.
Please note that for this step the existing installation is not changed at all, and all devices continue to run their legacy code and use their legacy communication.
For the Enterprise System to be able to understand the Data that is made available, digital twins of the Modbus deployment should be realized.

Please also note that the Modbus sniffer should not just be a dumb device, as this would introduce additional security risks. Instead, we propose an intelligent device that can publish named Data, e.g. using MQTT[2], and that also encrypts all Data it is transmitting using state of the art encryption technologies like TLS[3]. In our case, this was realized with a so-called periNODE[4]. The cabling be realized using Ethernet , in our case Single Pair Ethernet(SPE)[5] cables. In our opinion, this is the minimal approach that is necessary to become compliant with European law.

The major advantage of this minimal approach is that the Modbus installation can remain unchanged, the Modbus applications on the devices can remain unchanged and still IoT Data becomes available. Using this IoT Data we can now respond to the customers' requests for product relevant Data in accordance to the Data Act of the European Union.

Figure 2 visualizes the minimal approach.

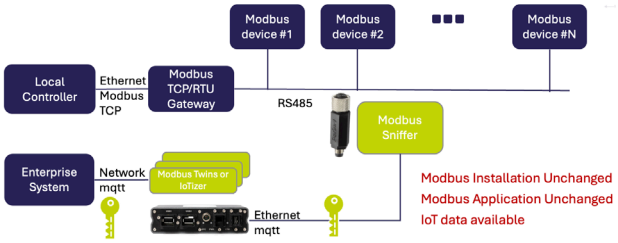


Figure 2 IoT Add On Data Sniffer

2.2 Introducing Secure Field Communication

While the minimal approach suffices to be compliant with the Data Act of the European Union, the Modbus itself remains insecure. Therefore, in order to remove the security risks mentioned above, (parts of) the RS-485 cabling shall be replaced with SPE. To enable legacy devices to use SPE, we once more use a periNODE, similar to the way we previously used a sniffer on the Modbus. However, instead of having only one in the whole system, we now connect one of them to each Modbus device. This way, the Modbus devices remain unchanged and the applications running on the Modbus devices also remain unchanged.

On the security side, however, we now have a secure communication between all the devices, as all of the periNODEs realize Modbus TCP over SPE by transmitting TLS encrypted messages among each other.

Therefore, even if anyone from outside would gain physical access to the cables and install a sniffer, they would no longer be able to interpret the gathered Data.

The big advantage of this step is that devices and applications remain unchanged, and while we still use Modbus communication, it is now secure as it is tunneled over SPE.

Figure 3 visualizes the new, secure communication.

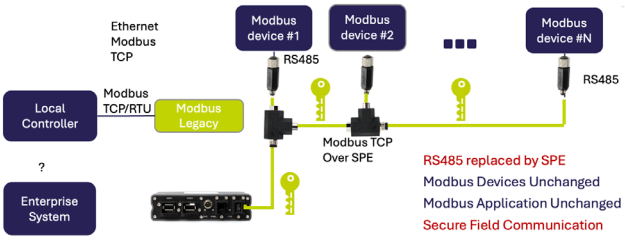


Figure 3 Switching from RS485 to SPE

2.3 Connecting the Enterprise Systems using Linux Containers

In the next step we want to connect all of our newly realized secure communication over SPE with the enterprise system. Of course, we do not want to re-cable our enterprise system with SPE. Therefore, we use our small industry PC to run the digital twin of our Modbus system, and also to connect to conventional Ethernet. This way, when using the enterprise system, we only need to connect to the digital twin and can retrieve all data we are interested in. This is shown in Figure 4.

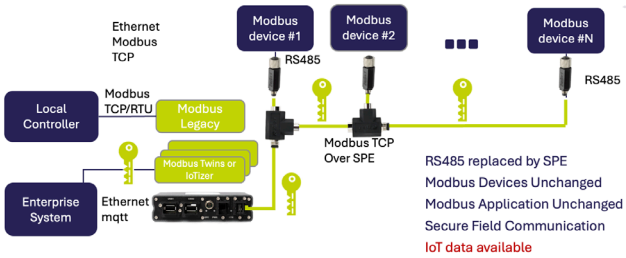


Figure 4 Adding IoT Connection

2.4 Updating Field Devices to SPE

Now that we have full access from the enterprise system, we can safely start replacing some of our devices. Of course, we do not want to replace all of them, especially not at once as stated before. However, there might be some devices that have broken down partially or deliver only reduced functionality for whatever reason. If these

devices are quite old, there often are no replacement parts necessary for repair available anymore. In this case, replacing to old device with a new Modbus TCP device can increase productivity. On the other hand there might be devices which still work, or for which there are simply no new versions available. The big advantage of this step by step approach is that we can have a mixed system – some old devices with added secure communications that still run the old code working together with newly installed devices. And due to the added bandwidth and scalability that result from the use of SPE instead of RS-485, it is even possible to keep all of the old devices and just expand the system.
Figure 5 shows our running example, now with one of the old devices replaced by a new Modbus TCP device.

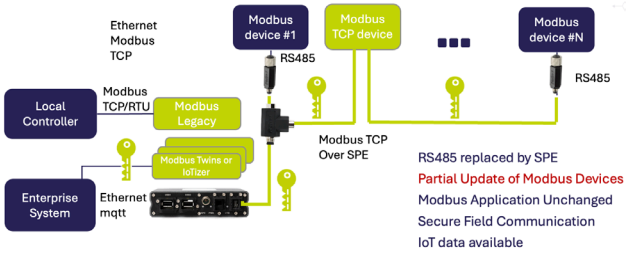


Figure 5 Updating Field Devices to SPE

2.5 Changing Communication to IoT communication

Currently, we are having powerful SPE cabling but only using it to transmit Modbus TCP data. In this step, we want to take things one step further and introduce IoT communication. The typical protocol used in new deployments nowadays is MQTT as it has numerous advantages over Modbus. Therefore, in this step, we switch to MQTT/HTTPS over SPE. As we still have our application running as Modbus client software, we introduce a virtual Modbus between it and our SPE, which runs on the industrial PC, in our case a periMICA[6].

However, even though we update our communication to MQTT, legacy devices can still be used with the adapters, as these can translate between Modbus and MQTT directly at these devices.

One huge advantage that is introduced here is that new devices that are added to our deployment now can either be Modbus devices or new smart devices that are incompatible with Modbus but can communicate using MQTT, resulting in IoTized Plug'n'Play devices.

Another advantage in this that there is no need to change any hardware – the whole step is only a software update.

Figure 6 shows our running example with a new deployment layout. The virtual Modbus and a smart device have been added.

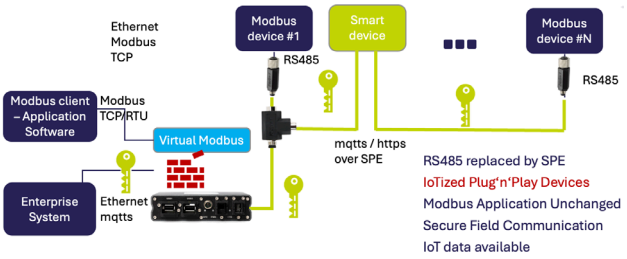


Figure 6 Changing to IoT Communication

2.6 Final Step: Introducing new application Software

In this last step we gradually replace the old application Software with a new one that uses secure IoT Protocols and the additional features those Protocols provide.

Figure 7 shows our running example again, now with a completely IoTized application software and another smart device added.

However, the legacy device that still can be seen will remain, as it is still working according to specification and fulfilling all current requirements. For this to work, we also keep one of the adapters in the system, which will continue to translate between MQTT/HTTPS and Modbus for that legacy device until the legacy device breaks or becomes obsolete.

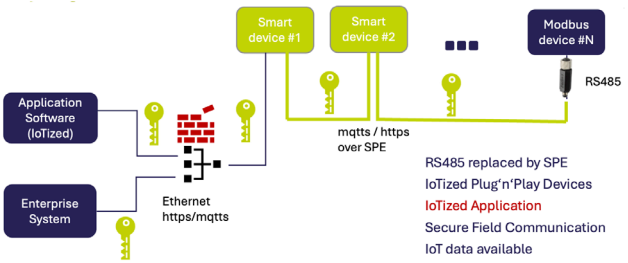


Figure 7 Transition from Modbus to secure IoT finished

3 Conclusion

In this paper we have shown a step-by-step approach for replacing Modbus installments with secure IoT installations as is required in order to confirm with upcoming EU regulations. This can also be seen as a chance to introduce additional security measures.
The big advantage of this step-by-step approach is that the deployment can continue to operate during the changes, we introduce minimal to zero downtime. Also, legacy devices that cannot be repaired anymore can now be replaced with modern smart devices. On the other hand, legacy devices which are still operational do not have to be changed as a smart adapter was introduced that can connect to their Modbus in-/output and translate between Modbus and an IoT protocol like MQTT.

DOI 10.5162/iCCC2026/P35

4 Bibliography

[1] Data Act of the European Union https://digital-strategy.ec.europa.eu/en/policies/data-act last accessed 2025-10-24

[2] MQTT https://mqtt.org/ last accessed 2025-10-24

[3] TLS https://datatracker.ietf.org/doc/html/rfc8446

[4] periNODE Modbus https://www.perinet.io/de/produkte/smart-components/perinode-modbus last accessed 2025-10-24

[5] Single Pair Ethernet according to IEC 63171-6:2021 https://www.vde-verlag.de/iec-standards/250428/iec-63171-6-2021.html

[6] periMICA https://www.perinet.io/de/produkte/edge-computer/perimica last accessed 2025-10-24

DOI 10.5162/iCCC2026/P36

Performance Assessment of 5G for Remote Sensing: KPI Measurements and Practical Insights

Keyvan Shahin, Christian Herglotz, Brandenburgische Technische Universität Cottbus-Senftenberg, Cottbus, Germany
keyvan.shahin@b-tu.de, christian.herglotz@b-tu.de

Abstract

Accurate performance measurements are essential for evaluating how well a 5G network can support practical sensing and data-intensive applications. To support the coverage study, we first developed a measurement framework for collecting key performance indicators (KPIs) of the 5G deployment. The framework, implemented using iperf3, was designed to provide reproducible measurements of throughput, latency, jitter, and retransmissions under a variety of configurations. This foundation enabled the collection of performance data during mobile trials and provided the basis for the coverage mapping and advanced analysis presented in this paper.

1 Introduction

Massive numbers of devices can be supported by fifth-generation (5G) cellular networks, which promise multi-gigabit throughput and sub-10 ms latency [10]. 5G is appealing for data-intensive IoT and sensing applications because of these features. For instance, ubiquitous sensing with low latency and high data rates is essential to environmental monitoring systems and smart city systems. 5G can be used for real-time processing of high-resolution sensor data on edge servers for applications such as unmanned aerial vehicles (UAVs) [1]. Theoretically, 5G's high bandwidth and low latency significantly broaden the range of real-time sensing applications by enabling to perform tasks in which 4G fails, like real-time analytics and 4K/8K video [2].
However, these promises might not be fully realized in actual 5G deployments, particularly in the early releases and implementations. Many operators first implement 5G in non-standalone (NSA) mode on legacy hardware, and coverage is still only partially available. According to [3], new applications will frequently see a mix of 4G and 5G connections. Standard speed tests or simulations may overestimate real-life performance in these circumstances. For IoT/sensor workloads, it is therefore essential to measure end-to-end KPIs under realistic circumstances, particularly throughput and latency. We can only comprehend whether 5G can satisfy the requirements of remote sensing and how deployment factors (location, mobility, and time of day) impact performance with the help of such data.
In this paper, we provide a practical measurement framework and performed an actual field evaluation of these tools for 5G in a campus sensor network setting. Following the methodology explained in [5], we install Raspberry Pi devices with 5G modems and use iperf3 [11] to automate tests [4,5]. After that, we carry out a number of tests to evaluate the uplink and downlink performance of the campus 5G network under actual loads. Quantifying the achievable throughput and latency, identifying variations over time and space, and drawing conclusions for 5G remote sensing applications are the objectives. These basic measurements are primarily performed to validate the functionality of our measurement setup. The main contributions of this paper are:

- Measurement framework: Using Raspberry Pi nodes and iperf3, we create an automated test platform for repeatable 5G throughput and latency measurements.
- Field evaluation: We conduct a study on campus to measure the latency and end-to-end throughput of a commercial 5G deployment under different traffic scenarios similar to that of the Internet of Things.
- Analysis: To highlight the advantages and disadvantages of the 5G network for sensor data, we examine the effects of various factors on the observed KPIs.
- Useful insights: In light of our findings, we provide recommendations for efficient 5G configurations.

This is how the remainder of the paper is structured. The relevant work on 5G KPI measurements and sensing applications is reviewed in Section 2. Our experimental setup and methodology are described in Section 3. The measurement results and discussion are shown in Section 4. A summary of the findings and recommendations for the future are provided in Section 5.

2 Related Work

The goal of recent field research has been to measure the practical performance of 5G networks. One of the first extensive 5G measurements was carried out by [5], who recorded physical-layer metrics and application throughput using commercial smartphones and specialized tools. Although 5G links can reach gigabit speeds under ideal circumstances, they discovered that latency was higher than intended for ultra-low-latency applications and that transport-layer inefficiencies (such as TCP (Transmission Control Protocol)) limited end-to-end utilization. [6] conducted citywide measurements of low- and mid-band 5G vs. 4G in their more recent work. They state that current 5G downlink throughput, with good coverage, is significantly higher than 4G LTE, averaging about 700 Mbps (peaking at about 1 Gbps). The authors point out that larger channel bandwidths, not new modulation or MIMO layers, account for a large portion of this gain.
The authors of [8] focus on methods for extracting KPIs from 5G equipment. After comparing commercial scanners, probe-based tools, and AT-command interfaces, they come to the conclusion that getting fine-grained real-time

DOI 10.5162/iCCC2026/P36

KPI data is difficult and frequently necessitates expensive setups. Their research emphasizes the necessity of useful measurement frameworks, such as the one we proposed in this work.

In [7] the authors investigated sensor-equipped UAVs with 5G capability. Their tests demonstrate that uplink latency stays bounded and the application operates as intended when the 5G link is not overloaded. However, pushing the UAV's data rate beyond the 5G capacity caused increased latency, dropped packets, and degraded performance. These studies highlight the importance of throughput and latency as key performance indicators (KPIs) for 5G-based sensing platforms and the necessity of conducting practical tests to determine their limitations.

Finally, a variety of testbeds and tools have been used for 5G experiments. Dedicated campus or private networks allow controlled evaluation: e.g., the Campus5G project deployed a full campus-wide private 5G network (with open RAN (Radio Access Network) architecture) to enable experimental research. Common practice is to use commodity hardware like Raspberry Pis as user equipment, and standard utilities like iperf3 to generate traffic. Such tools provide reproducible measures of throughput and latency (as also noted by industry guides). In our work, we similarly leverage Raspberry Pi endpoints running python scripts using iperf3 library to collect KPI data in a real 5G deployment, combining lessons from these prior testbeds and measurement studies.

3 5G Testbed Configuration and Measurement Methodology

This chapter details our 5G network testbed's hardware and experimental configuration, including the arrangement of the Raspberry Pi devices and the configuration and execution of iperf3-based measurements.

3.1 Hardware Setup and Deployment

Four Raspberry Pi 4 Model B boards were used to construct the testbed, and each board was linked to a Waveshare USB–M.2 (B-key) adapter that was carrying a 5G modem module.

In our configuration, one Pi functioned as the server and three Pis as clients that generated traffic. Our Python scripts, which used the iperf3 library, and Raspberry Pi OS 13 were installed on every device in order to automatically mate traffic tests.

In **Figure 1**, the position of the 5G antenna of the base station and the position of the Pis are illustrated by red and yellow markers, respectively. The client modems function in non-line-of-sight conditions because buildings obstruct the line-of-sight (LoS) between our office and the antenna. This makes an average signal condition to study the 5G network performance. We positioned each USB dongle (with its external antennas) directly outside a window to enhance signal reception. The server was situated approximately two meters away from the three clients which were arranged side by side at one window (**Figure 2**).

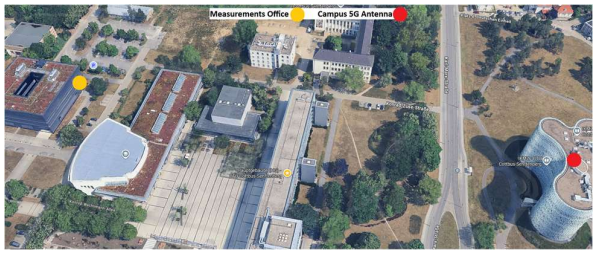


Figure 1 Campus antenna and measurement site

Figure 2 Client 5G dongles setup

3.2 Software Tools and Measurement Scripts

All traffic generation and measurement were done with the iperf3 tool (version 3.x). We developed a multiport server and a client Python script using iperf3 library. Using these scripts, we are able to conduct measurements with desired configurations and log the KPIs throughout these tests.

3.3 Measurement Configuration

Each test run is controlled by a set of input parameters. We repeat each experiment multiple times to gather statistics. The main configuration parameters are:

- Protocol: Either TCP (default) or UDP (User Datagram protocol) for traffic.
- Port: the TCP/UDP port used for iPerf3.
- Duration: 10 (default), seconds per run (the time interval over which iPerf sends data).
- Bandwidth: infinite (uncapped, default) (e.g. 10M for 10 Mbps).
- Parallel Streams: 4 (default), the number of simultaneous iperf3 streams per client).
- Number of Runs: 10 (default), total repetitions of the test, to allow averaging.
- Block Size (TCP window): 131072 (default) bytes, this is the iPerf3 default buffer size.

These configuration options, which are constant throughout each test, are set in the Python scripts. To investigate

DOI 10.5162/iCCC2026/P36

their impact on performance metrics, we changed one or more of these configurations to make various test scenarios.

3.4 Performance Metrics

From each iperf3 test we log the following KPIs (as reported by iperf3) along with a timestamp:

- Bytes Sent: Total data transferred by the client.
- Throughput (Mbps): Achieved data rate in megabits/second.
- RTT (ms): Round-Trip Time measured by iperf3 on the TCP connection (computed as the average TCP ACK latency).
- Retransmits: Number of TCP retransmitted segments (for TCP tests only).
- Jitter (ms): Packet delay variation (measured during UDP tests).
- Packet Loss (%): UDP packet loss percentage (UDP tests only).
- CPU Utilization (%): The operating system CPU load on each Raspberry Pi (logged via a system call on the Pi).

Additionally, we timestamp each run to later correlate with the logs from other server/clients.

3.5 CPU Usage and Power Considerations

During tests, we kept an eye on each Pi's CPU utilization as an indicator of the power consumption for Raspberry Pi devices [9].
Overall, by monitoring traffic parameters and observing how they affect throughput, latency, and other performance metrics, this testbed and measurement methodology enable us to systematically study 5G network KPIs. To isolate the effects of each setting on the recorded KPIs, the above detailed input configurations were changed across scenarios.

4 Measurements and Results

We carried out controlled experiments with different channel bandwidth, bulk transfer size, transmission duration, and client concurrency to investigate how configuration decisions impact KPI behavior.

Figure 3 Impact of Bandwidth on throughput

First, we discuss the networks performance in terms of throughput. Network performance was clearly impacted by changing the configured bandwidth. Restrictive bandwidth values resulted in lower throughput (**Figure 3**) and slightly higher latency (**Figure 4**). These patterns demonstrate that both achievable data rate and latency are affected by available bandwidth.

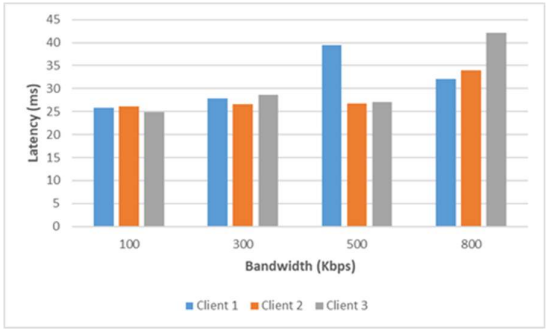


Figure 4 Impact of Bandwidth on latency

Figure 5 Impact of Bulk Size on Throughput

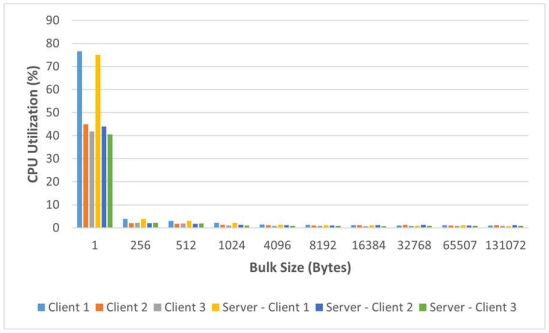


Figure 6 Impact of Bulk Size on CPU Utilization

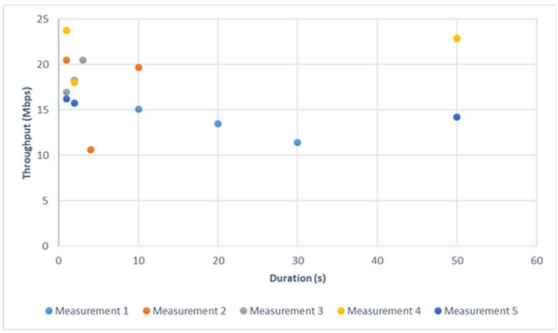


Figure 7 Impact of Transmission Duration on Throughput

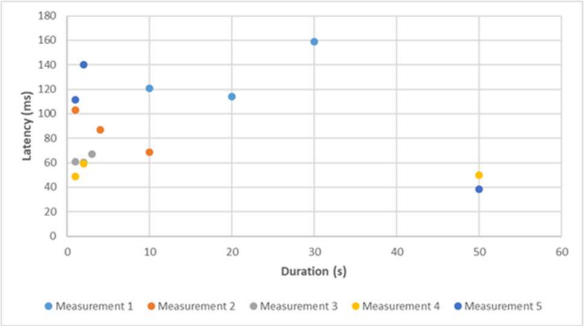


Figure 8 Impact of Transmission Duration on Latency

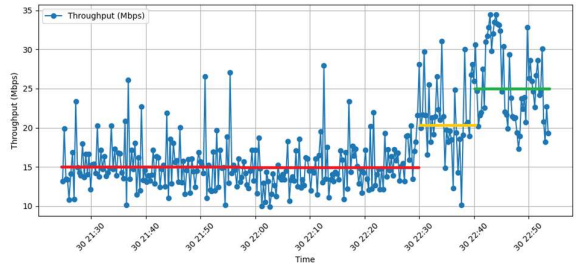


Figure 9 Impact of Interference on Throughput

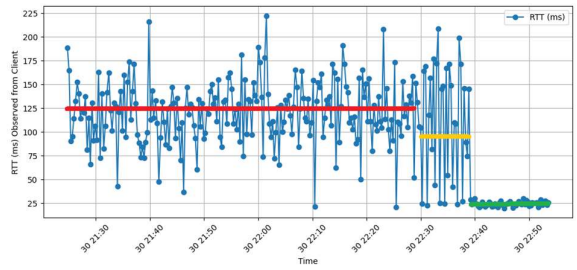


Figure 10 Impact of Interference on Latency

Performance was not noticeably impacted by bulk transfer size unless the bulk size was set too low (**Figure 5**). While values over 256 Bytes contributed to low CPU utilization, very small bulks introduce very high CPU utilization (**Figure 6**).

Figure 7 and **Figure 8** indicate that the duration of transmission seems to have no meaningful impact on the throughput and the latency.

We also ran multiple clients with different end times to investigate interference. As each client's task was finished, we clearly observed an increase in the throughput and a decrease in the latency (round-trip time) of the other working clients. We are demonstrating the throughput and latency of the last client to show how they change when other clients' tasks are finished (**Figure 9**,**Figure 10**).

We used UDP to repeat the majority of the tests, but most UDP runs failed or timed out in our non-line-of-sight weak-signal environment. Additionally, there were significant differences in CPU usage: TCP rarely exceeded ~10%, while UDP saturated almost 100% of the server CPU. Therefore, TCP proved to be more reliable and much less demanding for our sensing platform, even though UDP can theoretically deliver higher throughput.

5 Conclusion and Future Work

In this work, we developed an automated tool to measure 5G network KPIs under various configurations and real-world scenarios. We then share some of our findings, considering the impact of those configurations on the KPIs here.

As a continuation of this work, the developed tools and methods are being used in conjunction with drones and land vehicles to create a 5G network coverage map of the campus.

6 References

[1] K. Shahin et al., "A Modular Communication Architecture for Adaptive UAV Swarms," *2023 IEEE International Conference on Omni-layer Intelligent Systems (COINS)*, Berlin, Germany, 2023, pp. 1-6, doi: 10.1109/COINS57856.2023.10189245.

[2] C. Herglotz et al., "Designing Intelligent Sensor Networks: A Comprehensive Survey of Sensing, Edge AI, and 5G+/6G Connectivity," in IEEE Access, vol. 13, pp. 172306-172325, 2025, doi: 10.1109/ACCESS.2025.3615205.

[3] Banerjee, A. et al. (2024), "5G enabled smart cities: A real-world evaluation and analysis of 5G using a pilot smart city application," Internet of Things, vol. 28, p. 101326.

[4] Gkagkas, G. et al. (2025), "The Behavior of an IoT Sensor Monitoring System Using a 5G Network and Its Challenges in 6G Networking," Electronics, vol. 14, no. 16, p. 3167.

[5] Xu, D. et al. (2020), "Understanding Operational 5G: A First Measurement Study on Its Coverage, Performance and Energy Consumption," Proc. ACM SIGCOMM '20. Pages 479 – 494 https://doi.org/10.1145/3387514.3405882

[6] Rochman, M. I. et al. (2024), "A Comprehensive Real-World Evaluation of 5G Improvements over 4G in Low- and Mid-Bands," arXiv preprint arXiv:2312.00957.

[7] Damigos, G. et al. (2023), "Performance of Sensor Data Process Offloading on 5G-Enabled UAVs," Sensors, vol. 23, no. 2, p. 864.

[8] Kesir, S. et al. (2024), "A Comparative and Measurement-Based Study on Real-Time Network KPI Extraction Methods for 5G and Beyond Applications," arXiv preprint arXiv:2504.16039.

[9] Kesrouani et al., A Preliminary Study of the Energy Impact of Software in Raspberry Pi devices. 29th IEEE International Conference on Enabling Technologies: Infrastructure for Collaborative Enterprises, 2020, Bayonne, France. hal-02936861

[10] Attar H et al., 5G System Overview for Ongoing Smart Applications: Structure, Requirements, and Specifications. Comput Intell Neurosci. 2022 Oct 11;2022:2476841. doi: 10.1155/2022/2476841. PMID: 36268153; PMCID: PMC9578857.

[11] https://iperf.fr/iperf-doc.php

DOI 10.5162/iCCC2026/P37

Mobility-Induced Sensitivity of UAV-based Nodes to Jamming in Private 5G Airfield Networks – An Experimental Study

Pavlo Mykytyn[1], Ronald Chitauro[2], Onur Yener[1,2], Peter Langendoerfer[1,2]
[1]BTU Cottbus-Senftenberg, Cottbus, Germany
[2]IHP- Leibniz Institute for High Performance Microelectronics
mykytyn@b-tu.de

Abstract

This work presents an experimental performance evaluation of a private 5G airfield network under controlled directional SDR jamming attacks targeting UAV-based UE nodes. Using a QualiPoc Android UE, mounted as a payload on a quadcopter UAV, we conducted a series of experiments to evaluate signal degradation, handover performance, and service stability in the presence of constant directional jamming. The conducted experiments aimed to examine the effects of varying travel speeds, altitudes, and moving patterns of a UAV-based UE to record and analyze the key physical-layer and network-layer metrics such as CQI, MCS, RSRP, SINR, BLER, Net PDSCH Throughput and RLF. The results of this work describe the link stability and signal degradation dependencies, caused by the level of mobility of the UAV-based UE nodes during autonomous and automatic operation in private 5G Airfield networks.

1 Introduction

Fifth generation (5G) networks, including private 5G Campus networks deployed at critical infrastructures like airfields, promise high throughput and ultra-reliable connectivity. Their deployment is becoming more dominant, as organizations try to leverage and benefit from designing and operating their own 5G solutions. The deployment of these networks is driven by the need for enhanced security, reliability, and performance in industrial and academic environments [1]. However, these networks remain susceptible to targeted radio frequency (RF) jamming attacks, enabled by low-cost Software Defined Radios (SDRs), where a malicious attacker emits interference to degrade the signal-to-noise ratio (SNR) of legitimate communication links. Recent research works such as [2], present new highly sophisticated jamming techniques that selectively target certain portions of carrier bandwidth to maximize disruption while minimizing the jammer's power and detectability. In particular, the Synchronization Signal Block (SSB) that carries the primary and secondary synchronization signals and Physical Broadcast Channel (PBCH) with critical system information are known weak points. Attacks on the SSB can block user equipment (UE) from accessing periodic synchronization, effectively causing denial-of-service on the cell [3]. On the other hand, with a powerful transmitter, full-band barrage jamming, can still be used to overwhelm the 5G downlink.

Unmanned aerial vehicles (UAVs) introduce new dimensions to this threat landscape. UAVs can serve both as tools for network monitoring and as potential jamming platforms. Prior works, such as [4], have even proposed mounting low-power jammers on small drones to hover near base stations and disrupt 5G signals in a stealth manner. At the same time, equipping drones with cellular test equipment allows researchers to probe network performance in the air and prepare for the deployment of UAVs as the 5G Nodes. However, high node mobility in 5G networks poses additional challenges, such as problematical handover procedures and shortened connection time within single cell tower range. Previous works, such as [5], have even attributed the handover failures in 5G networks to factors such as high UE mobility, intentional interference, and signal unreachability.

The integration of UAVs with private 5G networks at airfields and other areas with limited access, promises enhanced connectivity and high data throughput rates to improve user experience, security, and area management applications [6]. However, the high mobility of UAVs and their exposure to adversaries raise critical concerns about link stability, especially at the network edge. Among the most critical threats is targeted jamming, which can disrupt both initial access and ongoing sessions, potentially compromising critical services of an airfield.

The interchange between high node mobility and jamming effectiveness is thus an important open question. While studies, such as [1], [6] offer valuable insights into the theoretical aspects of 5G security and related simulations, they do not involve experimental testing or validation of such use cases in the real-world scenarios.

Our research work aims to fill this gap by experimentally testing directional jamming attacks focused selectively at the SSB versus PBCH portions of the communication spectrums under normal communication conditions and cell tower handover scenarios using a constant barrage jammer. We evaluate the signal degradation and link stability using key physical-layer and link-layer metrics. The scope of this research work, however, is limited to experimental testing and evaluation of the influence of flight speed of UAV-based UEs on the effectiveness of a barrage jammer, targeting the PBCH segments of the 5G communication spectrum.

The rest of the paper is structured as follows: In Section 2 we present our experimental setup and methodology. In Section 3 we describe the obtained results, and in Section 4 we conclude this work.

2 Experimental Setup

The experiments were conducted on a private 5G Campus network deployed at a small fully operational airfield. The

DOI 10.5162/iCCC2026/P37

gNodeB of the base station and the network were configured for wide-area coverage of the airfield with two active cell towers operating in a sub-6 GHz 5G band.

2.1 UAV-based QualiPoc UE

A quadcopter UAV was fitted with a Rohde&Schwarz QualiPoc Android network quality testing device representing the UE to measure the 5G performance and collect physical-layer and link-layer key signal quality metrics in real time during an automatic mission flight as presented in the Figure 1.

To emulate realistic usage patterns and generate network load, two distinct traffic scenarios were developed during the UAV measurement campaigns. In the first scenario, a large downlink-dominant use case was simulated by initiating a download of a 10 GB CentOS image file, resulting in sustained high-volume downlink traffic. In the second scenario, a balanced uplink and downlink traffic profile was established by maintaining a live video conference call using GoogleMeet. These two scenarios enabled the evaluation of jamming and mobility effects under both asymmetric and symmetric traffic conditions.

The UAV was programmed to fly multiple automatic missions with a constant altitude of 11 m and varying speed of 3 m/s, 6 m/s, and 12 m/s, depending on the mission. In each conducted experiment, the UAV followed a constant preprogrammed mission flight path, covering an area of ~2 km^2 as depicted in the Figure 2.

Figure 1 Quadcopter UAV fitted with a Rohde&Schwarz QualiPoc Android network quality testing device.

2.2 Jamming Hardware and Software

A constant directional RF jammer was set up on the ground in the center of the area covered by the automatic flight mission as depicted in the Figure 3. The hardware utilized for jamming in the experiments consisted of a Linux-based Dell Latitude E5470 laptop running Ubuntu 22.04 and a programmable USRP B210 SDR from Ettus Research. The USRP B210 supports frequency coverage from 70 MHz to 6 GHz, with an instantaneous bandwidth of up to 56 MHz and a maximum output power between 10 and 17 dBm. Additionally, the jammer was equipped with a high-gain directional antenna, aimed into the area of UAV operation, and an external power amplifier.

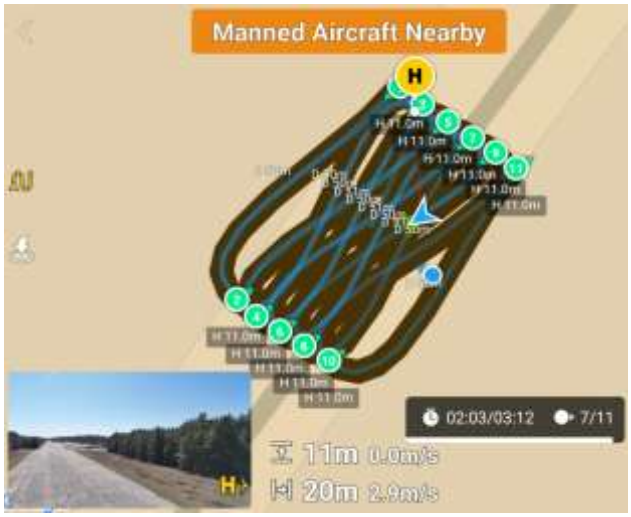


Figure 2 Preprogrammed automatic UAV flight mission path with a constant speed and altitude.

The transmission frequency was configured to target the SSB, in a partial-band jamming mode and PBCH portions of the 5G communication channel in the barrage jamming mode. The jammer transmitted a continuous pseudo-noise signal spanning a bandwidth between 10–40 MHz and transmission power between 20–30 dBm, depending on the targeted portion of the spectrum. To ensure the jammer was working correctly, during each experiment we recorded a spectrogram of the communication channel using a Rohde&Schwarz FSH8 spectrum analyzer.

Figure 3 SDR-based jamming hardware with a high-gain directional antenna and Rohde&Schwarz FSH8 spectrum analyzer.

DOI 10.5162/iCCC2026/P37

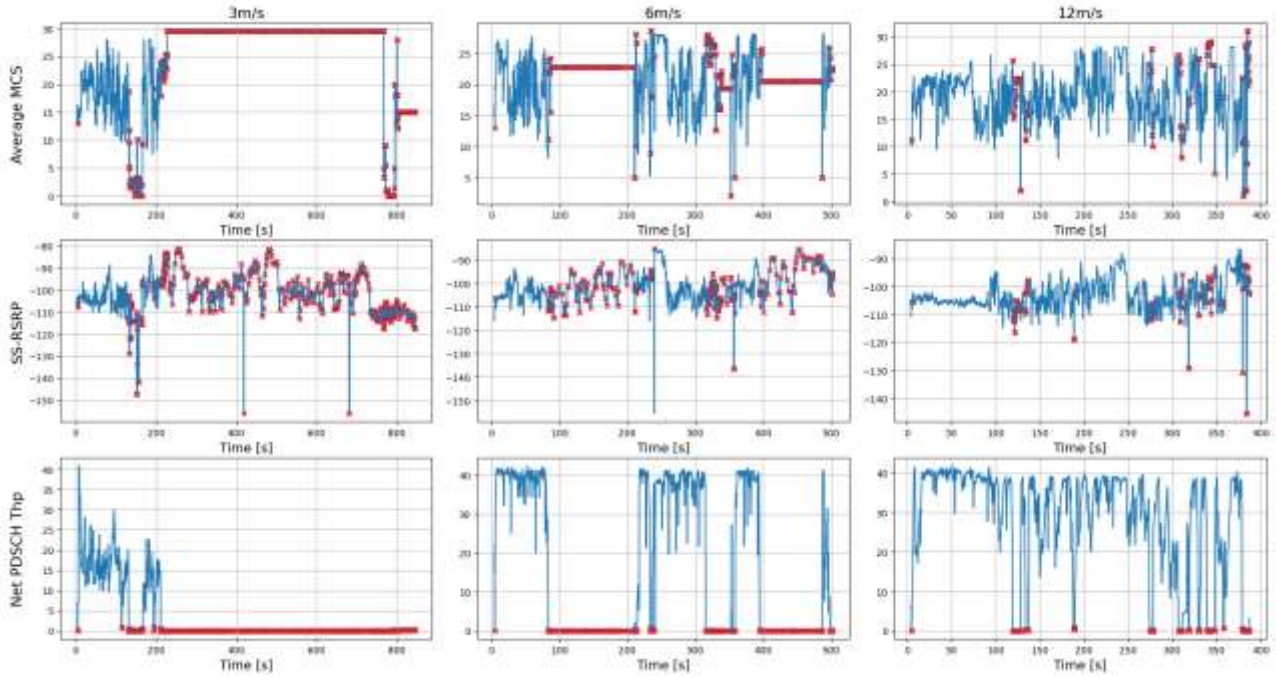


Figure 4 Comparison of the selected 5G channel quality metrics across different UAV travel speeds collected during the data link jamming campaign

3 Experimental Campaigns and Results

3.1 SSB Jamming Campaign

The first part of our jamming experiments, focused on partial-band jamming, targeted the portion of the 5G communication channel carrying the SSB frames. This measurement experiment was conducted on the ground before conducting the UAV flight missions. The targeted SSB bandwidth was 7.2 MHz wide, with the center frequency at 3743.10 MHz. The UE receiver was positioned 100 m away from the base station cell tower, and the jammer was placed at the mid-point between the UE and the cell tower, with its antenna aimed at the UE. During this measurement campaign, we initiated a download of a 10 GB CentOS image file, resulting in continuous downlink traffic. During this experiment we observed that jamming of the SSB portion of the spectrum was extremely successful. After jammer activation in each of the measurement sessions, the downlink traffic was completely disrupted, and the UE was disconnected from the base station. The throughput values, therefore, went down to zero.

3.2 Data Link Jamming Campaign

The full band jamming targeted the PBCH part of the 5G communication channel responsible for uplink and downlink communication. The SSB band was not affected during this jamming campaign. The targeted PBCH bandwidth was 60.8 MHz wide, with the center frequency of 3766.05 MHz. The jammer was configured to target a 20 MHz portion of the bandwidth and rapidly and repeatedly sweep through the 60.8 MHz frequency band to accomplish the jamming attack. The measurement sessions during this campaign were structured into 3 jamming experiments. The UAV was programmed to fly three separate identical missions at a constant altitude of 11 m and speed varying between 3 m/s, 6 m/s, and 12 m/s. To balance the uplink and downlink traffic during this experimental campaign, a live video conference call was initiated on the UE using GoogleMeet and maintained throughout all conducted experiments. The UE mounted to the UAV was programmed to collect relevant signal quality metrics and measure signal degradation under active data link jamming.

3.3 Signal Quality Metrics and Experimental Results

During the data link jamming campaign, we collected the Modulation and Coding Scheme Index (MCS), Reference Signal Received Power (RSRP), Signal-to-Interference-plus-Noise Ratio (SINR), Wideband Channel Quality Indicator (CQI WB), Block Error Rate (BLER), Retransmission count, and Net PDSCH Throughput values.

MCS is a unitless index representing the modulation type and coding rate applied to the physical downlink and uplink transmissions. Higher MCS reflects good channel conditions and lower MCS indicates fallback to low-throughput schemes under poor signal conditions. Monitoring MCS over time helps to evaluate the dynamic adaptability of the 5G link and its throughput efficiency. RSRP is measured in dBm and reflects the average received power of the synchronization signal, calculated at the UE. It indicates raw signal strength and is crucial for cell selection, handover, and beam management. Net PDSCH throughput reflects the actual UE data throughput received on the downlink, after accounting for retransmissions and protocol overhead and is a primary indicator of service quality.

Figure 4 represents the comparison between average MCS, RSRP, and PDSCH Throughput across different UAV speeds plotted over time. Each column corresponds to a fixed mission speed, while each row displays one of the three metrics. The red markers indicate the moments when downlink throughput fell below 1 Mbps, highlighting significant degradation of the communication link quality. At the speeds of 3 m/s and 6 m/s, the average MCS was higher than that of a 12 m/s, yet both 3 m/s and 6 m/s experiments experienced complete outages of the received signal caused by the jamming effects, while at 12 m/s, though with lower MCS values, the signal remained uninterrupted. While analyzing the RSRP across three different speeds, the values between experiments remained comparable, and did not show significant discrepancies, variating between - 90 and -110 dBm. This confirmed our expectations, because during the data link jamming campaign, only the PDSCH bandwidth was affected by jamming, leaving the SSB portion of the spectrum interference free.

While analyzing the Net Throughput across different speeds, the service quality variations between the speeds became even more clear. At 3 m/s and 6 m/s, the average recorded throughput was higher, however, the signal cut out in multiple places, with throughput dropping to absolute zero. Although the average throughput at 12 m/s was slightly lower, the connection always stayed active, avoiding complete service disruptions and thus, reducing the overall effect of the jamming signal on the link quality.

4 Conclusion

The results suggest that UAV mobility has an observable impact on the effectiveness of jamming attacks and 5G communication performance. Faster moving speed reduced the time the UAV spent inside high-interference zones, helping the UE escape jamming hotspots before link conditions could collapse.

However, it also had a negative effect on the 5G link stability and overall recorded throughput. Constant targeted jamming was very effective in all of the experiments, however, its impact on UE was most severe at lower traveling speeds. Higher UAV speeds, despite the lower average recorded MCS and throughout values, increased the resilience of the 5G link to targeted, directional and stationary jamming attacks by reducing the exposure time of UE to peak interference.

5 References

[1] M. Wen *et al.*, “Private 5G Networks: Concepts, Architectures, and Research Landscape,” *IEEE J. Sel. Top. Signal Process.*, vol. 16, no. 1, pp. 7–25, Jan. 2022, doi: 10.1109/JSTSP.2021.3137669.

[2] Y. Arjoune and S. Faruque, “Smart Jamming Attacks in 5G New Radio: A Review,” in *2020 10th Annual Computing and Communication Workshop and Conference (CCWC)*, Jan. 2020, pp. 1010–1015. doi: 10.1109/CCWC47524.2020.9031175.

[3] Mr. A. Birutis and A. Mykkeltveit, “Practical Jamming of a Commercial 5G Radio System at 3.6 GHz,” *Procedia Comput. Sci.*, vol. 205, pp. 58–67, Jan. 2022, doi: 10.1016/j.procs.2022.09.007.

[4] P. Skokowski *et al.*, “Jamming and jamming mitigation for selected 5G military scenarios,” *Procedia Comput. Sci.*, vol. 205, pp. 258–267, Jan. 2022, doi: 10.1016/j.procs.2022.09.027.

[5] A. Haghrah, M. P. Abdollahi, H. Azarhava, and J. M. Niya, “A survey on the handover management in 5G-NR cellular networks: aspects, approaches and challenges,” *EURASIP J. Wirel. Commun. Netw.*, vol. 2023, no. 1, p. 52, Dec. 2023, doi: 10.1186/s13638-023-02261-4.

[6] J. Cantizani-Estepa, P. Vera-Soto, C. Baena, E. Baena, R. Barco, and S. Fortes, “Analysis of 5G Technology in Airports: Requirements, Challenges, and Use Cases Evaluation,” *IEEE Access*, vol. 11, pp. 136109–136121, 2023, doi: 10.1109/ACCESS.2023.3336017.

DOI 10.5162/iCCC2026/P38

Further Development of a Measurement Method assessing the Coverage Quality of a private Standalone 5G network.

Johan Levon[1], Rolando Cortés Martinez[2], Ilkin Alkhasli[3], Keyvan Shahin[4]
[1] Chair of Automation, Brandenburg University of Technology Cottbus-Senftenberg (BTU C-S), 03046 Cottbus, Germany.
[2] Chair of Control Systems and Network Control Technology, BTU C-S, 03046 Cottbus, Germany.
[3] Fraunhofer Institute for Ceramic Technologies and Systems IKTS, Cognitive Material Diagnostics Group.
[4] Chair of Computer Engineering, BTU C-S, 03046 Cottbus, Germany.

Abstract

This paper demonstrates the methods that have been adopted and developed to measure the coverage quality of a private, standalone 5G network. Using the specific example of the private 5G network at the University of Cottbus-Senftenberg, this article presents the application of this method, from hardware to algorithms. The results show the relationships found between input factors such as variations in the physical context and in the network, and key performance indicators such as throughput and round-trip time.

1 Introduction

In recent years, the fifth-generation new radio (5G NR) mobile communication technology has been widely adopted across numerous industrial applications. Although there are some efforts to evaluate and characterize the status or quality of the network (see [1], [2]), evaluating a 5G stand-alone (SA) private network under the effects of variations in physical and network parameters remains to be explored. In particular, studies on mobile vehicles with 5G connectivity are limited (see [3]) because the current 5G public infrastructure is still limited. A further challenge in evaluating the performance of the 5G network for vehicles lies in the numerous input factors and their combinations. On the other hand, performance measurements involve selecting key performance indicators (KPIs) from a non-negligible list. For this reason, to obtain a sensible performance evaluation of the correlations between input factors and KPIs, the use of metadata analysis tools is crucial. The characterization of the 5G network also depends on the operator's configuration. In this work, we focus on the enhanced mobile broadband (eMBB) service, the most common service today [1].

The contribution of this work is the study of a 5G SA performance through a measurement campaign using 5G-enabled vehicles (on ground and in air) and the further analysis between relevant KPIs with respect to mobile characteristics like spatial position and velocities of vehicles, as well as the dependence under different data-related parameters like throughput and packet size.

2 Hardware used

For this Measurement Campaign a quadricopter unmanned aerial vehicle (UAV) and a cargo bike have been used as moving platforms. For our experiments we used the 5G SA private network provided at the Brandenburg University of Technology, Cottbus-Senftenberg (BTU-CS). The 5G network works indoors (in some Hallways) and outdoors. It supports the radio signals in the n78 frequency band with the eMBB service.

2.1 5G-Connected Devices

The measurement campaign is also designed to compare the performance of 5G SA-compatible hardware, either on mobile platforms or as a server. All devices are equipped with dedicated SIM cards and have the option of a fixed IP address. The SIM8200EA-M2 5G Hat from Waveshare is intended for a placement atop a Raspberry Pi 4 and based on the Qualcomm Snapdragon X55 platform. It offers compatibility with SA, NSA, 5G, 4G, and 3G, operates across multiple frequency bands including n78 (3.3 to 3.8 GHz), and supports eMBB. For more technical information, see [4] Another model, also from Waveshare, features the newer RM520N-GL 5G module, available as a hat for a Raspberry Pi 4 or as a dongle [5]. The Samsung Galaxy Tab S9 5G is included in the campaign for its broad network compatibility, specifically its support for 5G SA n78, although detailed information on its integrated 5G module is not provided [6]. The Telit FN990 data card, a compact board with an M.2 connection, is also utilized on the UAV; it supports 5G sub-6 GHz, SA and NSA operations, connects to the 5G core network, and is also compatible with band n78 [7].

2.2 5G-enabled UAV

The Telit FN990 data card is connected to the Lattepanda 3Delta processor, which has a 4-core, 11th generation processor running up to 2.9 GHz and 8 GB of RAM. The Lattepanda and FN990 antennas are attached to the rear of the UAV as a payload. The UAV is based on the X500 V2 frame and uses a Pixhawk 6C as its flight control unit. An additional power converter supplies the Lattepanda with 12V. Additionally, the Lattepanda 3Delta, which uses Ubuntu 22.04, is connected to the flight controller via the RS232 port for reading the UAV GPS and sensors, (see **Figure 1**).

2.3 5G-enabled Cargo Bike

The Raspberry and 5G Hat casing is mounted on the left side of the Cargo Bike with removable attachments, using a banded metal sheet to pinch the rim, (See **Figure 2**).

DOI 10.5162/iCCC2026/P38

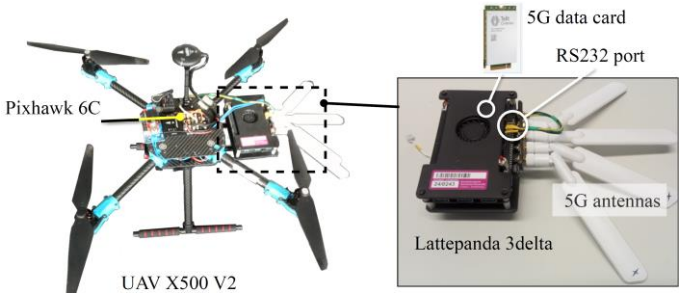


Figure 1 UAV adaptation with 5G data card. Author: Rolando Cortés-Martinez.

Figure 2 Driver point of view of the 5G-Hardware installed on the Cargo Bike. Author: Johan Levon.

A touchscreen installed on the handlebars serves as a human-machine interface. The 4 Antennas of the 5G Hat are between 70 and 96cm above the Ground Level.

The Galaxy Tab S9 was fixed facing the sky with the back double sided taped to the right side of the sitting surface of the cargo Bike. The integrated Antennas are at a height of 60cm above ground level.

3 Testing Methodology

The GPS position and its Coordinated Universal Time (UTC) are recorded every second. Algorithms test the latency and throughput of the connection to the 5G Network every 3 to 10 Seconds also with UTC. Both the positioning and performance are combined with UTC Timestamp. At the end we get a Map Showing the Performance corresponding to the localization on the Campus.

For the UAV recordings, the flights are carried out manually in public areas maintaining low speed, less than 3m/s, and in supervised automatic mode under empty-from-people conditions in the sport fields in the BTU installations. A total of 10 flights were recorded with different mission objectives. We tested different flight fixed altitudes of 5 m, 10 m and 20 m, but also a variable altitude from 0 m to 80 m with a constant elevation velocity. Another considered variable is the velocity, which was established at 1 m/s, 3 m/s, 5 m/s and 10 m/s. Finally, the orientation of the UAV (yaw angle ψ) is constantly varying as the target way points are continuously updated. It is important to note that although for the missions definition we have fixed some variables, for the analysis of the data we use all the recorded information (e.g., UAV at ground level).

3.1 Monitoring the Position through GPS-Signal

For the Cargo Bike, although it is possible to replace one antenna with a GPS Module on the 5G Hat. It was decided to use the Tablett as GPS-GNSS Receptor for positioning. Giving the technology of the tablet, the proximity to Buildings and that the Cargo Bike was moving during the measurement, the expected Precision for the positioning lays between 3 and 10m.[8]. On the other side, an already integrated GPS module was used to measure the position in the UAV case. The accuracy of the position is higher with the drone, given his flying height of 10-80 meters above ground level, laying between 1 and 5 meters.

3.2 Measurements

A first data collection for coverage evaluation using the Round Trip Time (RTT) KPI was carried out with the Cargo Bike in combination with the two mounted 5G devices. Significant connectivity differences have been noted.[9] **Figure 3** shows the quality of the connection to the 5G network depending on the location.

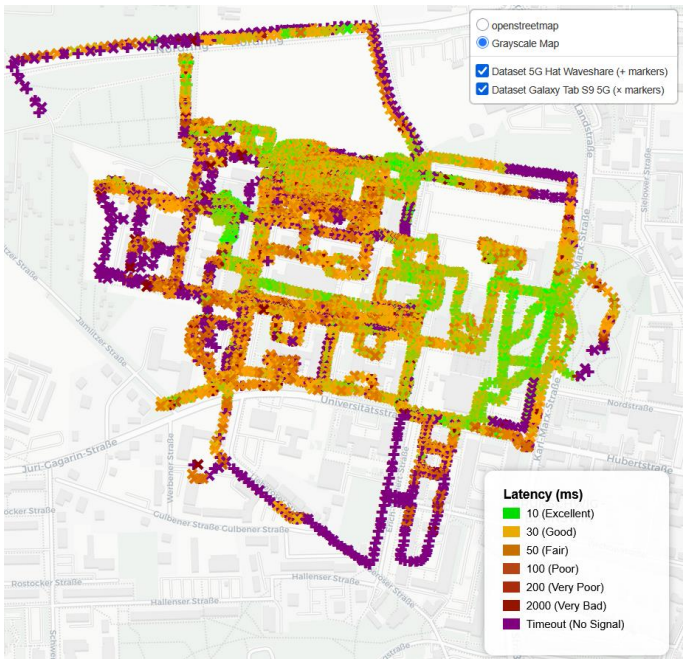


Figure 3 5G network coverage heat map with all the Data measured with the Cargo Bike. This interactive map is available on the iCampus Website [10]. Author: Johan Levon

To improve data collection, the iperf3 protocol was chosen. For simple ping tests, the fixed IP address of the 5G core router 192.168.88.100 is defined as the reference to ensure that all future measurements are comparable.

For the UAV the parameter and KPI measurements consists of the Reference Signal Received Power (RSRP), Reference Signal Received Quality (RSRQ), Signal to Noise Ratio (SNR), Yaw angle, the position of the UAV (latitude, longitude, and relative altitude), and velocities of the UAV (north velocity and east velocity). Additionally, the UAV captures the iperf3 measurements like the received data packages, throughput and RTT. For these tests we use the following configuration: TCP/IP protocol, port 5201 (single client connected to the server), 1 second *Duration* of the transmissions, 4 streams, infinity bandwidth, and 128 KB for the packet size. The latter value corresponds to the maximum size for TCP packages,

DOI 10.5162/iCCC2026/P38

which is suitable for obtaining maximum values for throughput and at the same time we test the network and buffer near to the limits. We perform variations in speeds and altitudes in the sport field test flights while in the rest of the campus area we maintain fixed secure conditions (see **Figure 4**).

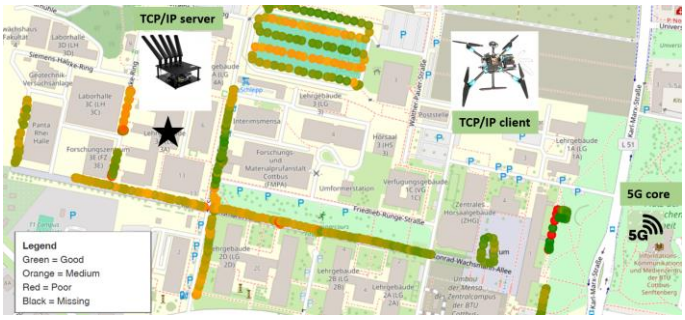


Figure 4 Campus 5G mapping for the UAV at 20m altitude.

As part of the flights carried out in the sport fields, **Figure 5** shows an exemplary extract of the flights carried out. In this example, the UAV performs a zigzag trajectory with fixed altitudes of 10m in t= (74, 77) min, another altitude of 20 m in t= (82, 84) min and a growing altitude up to 80 m in t= (85, 88) min. While the altitude is fixed at 10m or 20m the UAV surveys a series of way points at maximum nominal velocity of 3m/s and 5m/s, respectively. The directional angle yaw is also shown.

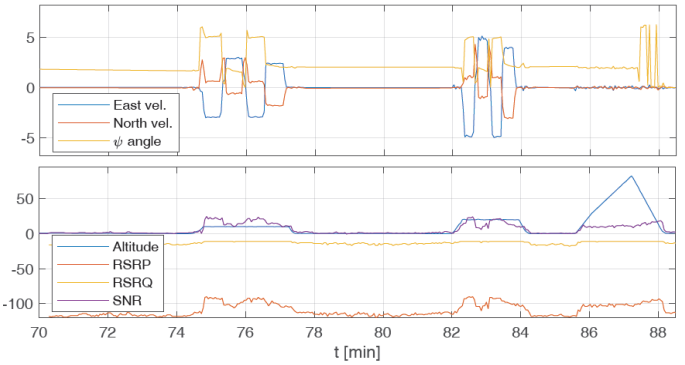


Figure 5 Extract UAV flight measurements campaign 2025.

4 Results and Data Analysis

This section describes the exploratory data analysis of the 5G signal measurements. The data consisted of two separate data sets, measured from the client and from the edge side simultaneously. As a first step, the two data sets were merged, synchronized based on the time stamps. Since the synchronization of the time stamps from the two measurements were not perfect, a 3 second tolerance was applied that assumed data points captured within the tolerance measure the identical signal occurrence. Upon closer investigation this assumption proved to be accurate.

The merged dataset contained missing values, which severely impacted the possibility to analyze it, this required the imputation of the dataset. Imputation is the process of estimating and filling in the missing values based on the available data. It is generally possible to produce a complete dataset without introducing significant bias or distortion.

Simple imputation methods such as filling missing values with the mean or the median of the column ignore the correlation between the features. In this study, the Iterative Imputer from scikit-learn was used, that provides a more sophisticate and multivariate approach. Here, each missing value of a feature was modelled as a function of the other features in a round-robin fashion. The iterative imputation process consists of the following steps:

- Initially the missing values are filled with the means of available values of that feature
- For each feature that has missing values, the remaining features are used to predict the missing values by using Bayesian Ridge regressor
- The process is continued for all features with missing values repeatedly, until a convergence is reached.

This method effectively uses the full feature covariance structure to produce more accurate estimates for missing values of throughput (Mbps), RTT (ms), CPU Util (%), RSRP, RSRQ, SNR, *position.lat*, *position.lon*, *position.alt*, *Direction* and *Bytes*.

In the last step, the Linear Generalized Additive Models (GAMs) was used to visualize the detailed correlation of separate variables with each other. Within the framework of this study, the *Throughput* was chosen as the main variable of interest, other variables were varied individually while keeping the rest constant to determine their impact on *Throughput*. GAMs are a powerful class of models that balance the interpretability of linear models with the flexibility to capture non-linear relationships. They are well-suited for analyzing complex, non-linear dynamics often observed in 5G network performance metrics.

A Linear GAM expresses the expected value of the target variable Y (in this case *Throughput*) as a sum of smooth, non-linear functions of the individual predictor variables (SNR, RTT, etc.), instead of a simple linear combination. Since relationships are additive across the functions in linear GAM, there are no interaction terms between the predictors, making the contribution of each feature easily discernible. Liner GAM has advantages for the 5G data analysis, such as interpretability, flexibility and robustness.

- Interpretability: Since the model is a sum of univariate functions, the impact of each feature on the target variable can be visualized and interpreted independently, providing clear insights into the non-linear operational dynamics of the 5G network.
- Flexibility: The smooth functions (in the internal components of GAM) allow the model to accurately fit complex, non-monotonic relationships between network parameters without requiring manual feature engineering.
- Robustness: They are generally more robust to outliers and misspecification than purely parametric models.

By using GAMs, the non-linear relationship between, for example, the RSRP and the resulting *Throughput* can be modeled precisely, offering a more realistic representation of network behavior than a simple linear regression. **Figure 6** shows the Partial Dependence Plots (PDP) of the network features on the *Throughput*, it consists of 10 different plots, each showing the individual impact of a network parameter on the *Throughput*. The X-axis represents the range of the predictor variable; the Y-axis represents the estimated contribution of the predictor variable (*Throughput* in this case).

DOI 10.5162/iCCC2026/P38

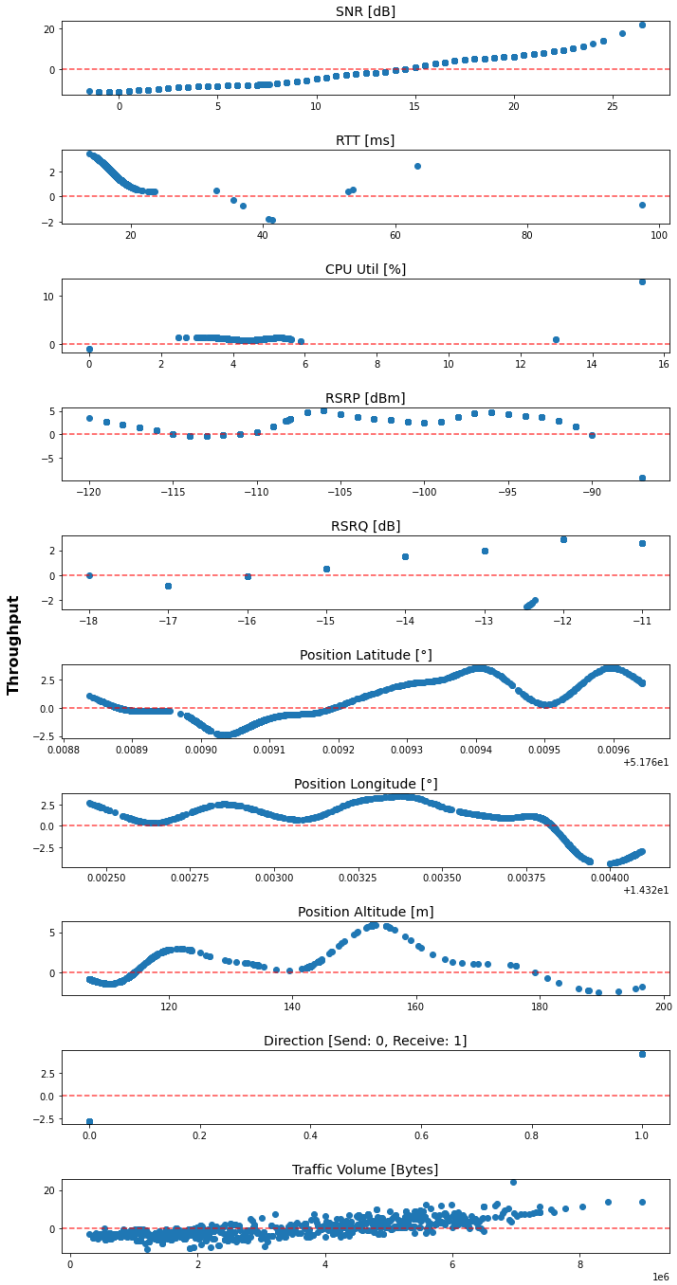


Figure 6 PDPs for Throughput showing the effect of the network features on the predicted network throughput, based on GAM analysis. In the Direction plot the value of 1 denotes Sent, a value of 0 Received.

From Figure 6, we can find some interesting relationships. For example, as RSRQ approaches -12.5 dB the *Throughput* exhibits a decay. A positive impact value indicates that the specific value of the predictor variable contributes to a higher predicted value of the target variable *Throughput*. A negative impact value indicates that the specific value of the predictor value contributes to a lower predicted value of *Throughput*. The steepness of the curve represents the sensitivity or magnitude of the effect at that point. The zero line is added to each plot to simplify the interpretation. The value is above, it improves the Throughput, below, it reduces it.

5 Perspectives

Further extension of the work can address the deeper analysis of the data through the extensive utilization of GAM modeling for the different KPIs versus input factors. Additionally, more PDPs should be used to show the correlations of more target variables. The current results can serve as a foundation to design more advance measurement campaigns focused on specific aspects of the 5G network, for example to analyses the effect of buildings and trees in the surroundings.

6 References

[1] Xu, D., Zhou, A., Zhang, X., Wang, G., Liu, X., An, C., Shi, Y., Liu, L., and Ma, H. Understanding operational 5G: A first measurement study on its coverage, performance and energy consumption. *In Proceedings*, 2020.

[2] Singh, B.C., Diaz, R., and Shetty, S. Evaluating the performance of 5G NR in indoor environments: An experimental study. *Procedia Computer Science*, 241, 2024, 16–23.

[3] Koumaras, H., Makropoulos, G., Batistatos, M., Kolometsos, S., Gogos, A., Xilouris, G., Sarlas, A., Kourtis, M., 5G-enabled UAVs with command and control software component at the edge for supporting energy efficient opportunistic networks, 2021.

[4] Waveshare, 2025, *SIM8200EA-M2 5G HAT technical characteristics,* lastly consulted on 01.12.2025, https://www.waveshare.com/sim8200ea-m2-5g-hat.htm

[5] Waveshare, 2025, *RM520N-GL 5G HAT,* 01.12.2025 https://www.waveshare.com/wiki/RM520N-GL_5G_HAT#Overview

[6] Lieske industry electronics, 2025, *Product information for Samsung SM-X716BZAAEUB*, 01.12.2025 https://industry-electronics.com/samsung/sm-x716bzaaeub-x716b-galaxy-tab-s9-11.0-5g-128-gb-lieske_2086953.htm

[7] Telit Cinterion, 2025, *FN990A40/A28 Product Features and Certification*, lastly consulted: 01.12.2025 https://www.telit.com/devices/fn990axx/

[8] Håkansson, M. Characterization of GNSS observations from a Nexus 9 Android tablet. GPS Solut 23, 21 (2019). https://doi.org/10.1007/s10291-018-0818-7

[9] Levon J., Ruffert C., Herglotz C., Nutzung privater 5G-Netzwerke in der Industrie: Entwicklung eines mobilen Demonstrators, um die Netzabdeckung zu testen. MikroSystemTechnik Kongress 2025 · 27. – 29. Oktober 2025, Page 430.

[10] iCampµs Cottbus, 2025, *5G-Connected-Sensing*, lastly consulted 01.12.2025, https://icampus-cottbus.de/5g-connected-sensing/

DOI 10.5162/iCCC2026/P39

Adaptive 5G Gateway for Smart Livestock Monitoring

Nina Pearl Doe[1], Hamid Zagariasl[1], Christian Herglotz[1]
[1]Brandenburg University of Technology Cottbus-Senftenberg, Germany
ninapearl.doe@b-tu.de, zargaria@b-tu.de, christian.herglotz@b-tu.de

Abstract

Internet of Things (IoT) systems in smart livestock farming requires resilient real-time or periodic monitoring of animal welfare and farm productivity. This paper presents an implementation of a 5G gateway (Raspberry Pi 5 and Quectel RM520N-GL) that bridges the farm-based anchor sensor mesh network to a cloud server, ensuring reliable cattle health data (pH, temperature, motion) transmission across varying network conditions. The proposed solution features an automatic network detection and adaptation between 5G Standalone (SA), 5G Non-Standalone (NSA), 4G (LTE), and Wi-Fi, a dual layer buffer management with Random Access Memory (RAM) and SQLite storage to ensure no data loss during network outages, and a priority-based message handling for critical health alerts. Comprehensive test results reveal mean latencies of 38.5ms (5G-NSA, 95th percentile: 44.9ms), 34.7ms (4G LTE, P95: 40.0ms), and 24.0ms (Wi-Fi, P95: 26.4ms), all with zero message loss. The gateway maintained stable operation across 47 dB signal range (-65 to -112 dBm), including private 5G-SA infrastructure during our multi-environment testing.

1 Introduction

Smart livestock farming requires continuous monitoring of animal health vitals including body temperature, pH, and activity levels [1]. While IoT sensors like bolus devices can continuously monitor vital parameters, another critical challenge lies in reliable data transmission from the farm locations to cloud-based analytics platforms. Fifth Generation (5G) cellular networks offer promising solutions with their low latency and support for massive Internet of Things (IoT) deployments [2]. Some farm locations (especially, rural areas) often lack consistent 5G coverage, experiencing network degradation or complete disconnection [3]. Some solutions typically rely on LoRaWAN connectivity, which can result in high latency and limited bandwidth for real-time monitoring of large herds [4].

This work presents a 5G gateway that bridges the gap between local sensor networks and cloud infrastructure and focuses on maintaining continuous operation through network adaptation and edge processing capabilities. We provide an empirical validation of the multi-network adaptive gateway architecture developed as part of the Smart Livestock Farming project at Brandenburg University of Technology (BTU) Cottbus-Senftenberg. Our contributions include: (1) A three-layer architecture with adaptive network selection, priority-based Quality of Service (QoS) management, and edge analytics, (2) Comprehensive empirical validation across 60-minute baseline tests on commercial 5G-NSA, 4G LTE, and Wi-Fi networks, (3) Demonstration of operation across a 47 dB signal range from excellent (-65 dBm) to marginal (-112 dBm) conditions, and (4) Dual-layer buffering achieving 100% message recovery during network outages.

2 Related Work

Previous agricultural IoT gateway implementations have primarily focused on single-network solutions. Talavera et al. [5] reviewed IoT applications in agriculture but noted the lack of multi-network resilience. Commercial solutions offer industrial reliability but are limited to LoRaWAN-to-4G bridging without 5G support [4].

2.1 5G for Smart Agriculture

The theoretical benefits of 5G for agriculture have been extensively studied in simulation. Liu et al. [7] discussed potential improvements in precision agriculture, while Meng et al. [8] analyzed low-latency benefits for real-time control.

The distinction between 5G Standalone (SA) and Non-Standalone (NSA) architectures is crucial for rural deployment. As explained by Alnaas and Alhodairy [9], 5G-NSA leverages existing 4G infrastructure as an anchor, making it more readily deployable but potentially compromising latency characteristics.

2.2 Network Resilience in IoT Systems

Resilient data transmission in challenged networks has been addressed through various buffering approaches. Jafri et al. [10] proposed multi-tier buffering for narrowband IoT, but agricultural environments require different optimization strategies due to extended outages and extreme signal variations. Our dual-layer approach extends this work with agricultural-specific adaptations.

3 System Architecture

Figure 1 depicts the system architecture showing the 5G gateway's position as the critical bridge between farm sensors and cloud infrastructure. The gateway consists of a Message Queuing Telemetry Transport (MQTT) broker, dual buffering, QoS management, edge analytics and adaptive network selection features.

The architecture consists of three layers with the gateway implementation as the main focus of our work:

- Farm Layer: Bolus sensors inside the cattle measure temperature, pH, and motion, transmitting data to anchor nodes distributed across grazing areas via Wi-Fi mesh network.

DOI 10.5162/iCCC2026/P39

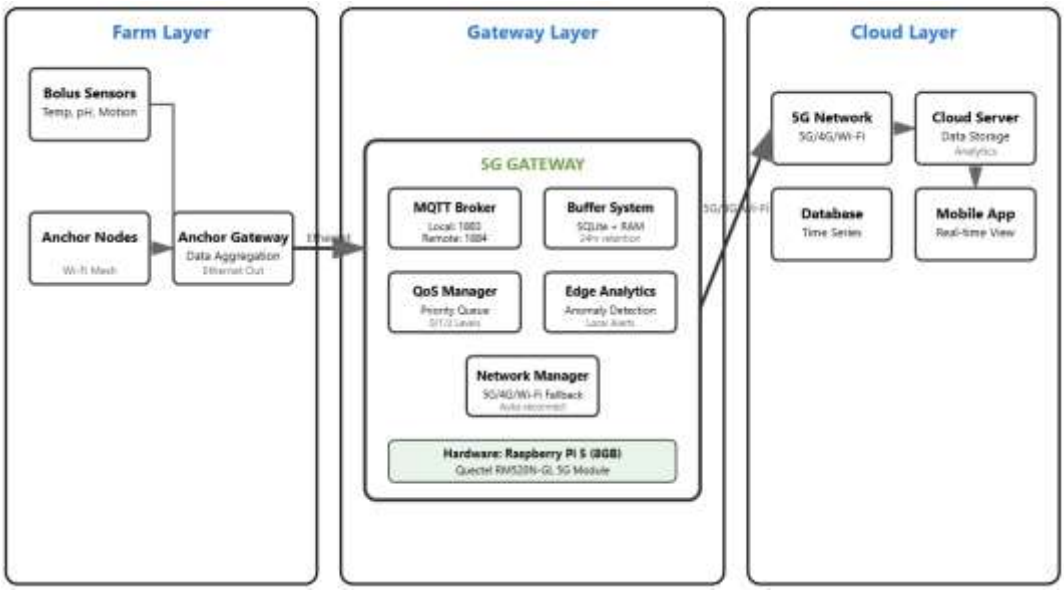


Figure 1 Proposed Sytem Architecture

- Gateway Layer: The adaptive 5G gateway receives aggregated data from the anchor node via Ethernet, implementing dual buffering, network selection, and edge analytics.
- Cloud Layer: Receives processed data for storage, analysis, and mobile app distribution.

3.1 Hardware Platform

The gateway hardware consists of a Raspberry Pi 5 (8 GB RAM)) with ARM Cortex-A76 processor at 2.4 GHz, providing sufficient computational capacity while maintaining low power consumption [11]. It is connected to a Quectel RM520N-GL 5G modem via USB-to-M.2 B-Key adapter. The RM520N-GL supports 5G New Radio (NR) Sub-6 GHz bands including n1, n3, n28, n41, n77, n78, and n79, as well as LTE bands B1, B3, B7, B8, B20, B28, B38, B40, and B41 [12]. The network status is monitored through Attention (AT) commands over a serial interface (/dev/ttyUSB2 at 115200 baud).

3.2 MQTT Bridge Architecture

The gateway implements a dual-client MQTT bridge [6]. A local Mosquito broker (port 1883) receives telemetry from the anchor node, while a dual-client system subscribes locally and publishes to the cloud broker (port 1884). This separation enables local buffering during cloud disconnection. All livestock/# topics are automatically forwarded with source validation and format checking, and ensures message persistence during cloud disconnections.

3.3 Quality of Service Manager

The QoS Manager implements priority-based message handling across three MQTT QoS levels. Critical health alerts such as fever detection or abnormal pH readings use QoS 2 (exactly one delivery) ensuring guaranteed transmission with duplicate message detection. Regular health telemetry including periodic temperature and activity readings use QoS 1 (at least one delivery) providing reliable transmission with potential duplicates. Location updates and non-critical status messages use QoS 0 (best effort delivery) minimizing overhead for high-frequency, loss-tolerant data. The edge analytics feature performs local anomaly detection for temperature deviations exceeding 2°C from baseline, generating immediate alerts without cloud dependency.

Network	Signal Threshold	Transmission Strategy	Message Batching
5G NR	> -70 dBm	Real-time streaming	None (immediate)
4G LTE (strong)	> -85 dBm	Light compression	10 messages
4G LTE (weak)	> -100 dBm	Heavy compression	50 messages
Wi-Fi (strong)	> -60 dBm	Real-time streaming	None (immediate)
Wi-Fi (moderate)	> -75 dBm	Light compression	10 messages
Wi-Fi (weak)	> -85 dBm	Heavy compression	50 messages
Offline	No signal	Local buffering only	Unlimited (to buffer)

Table 1: Adaptive Network Manager Thresholds and Strategies.

3.4 Adaptive Network Manager

The Network Manager continuously monitors signal quality through AT commands (including AT+QRSRP for cellular Reference Signal Received Power (RSRP), AT+QENG for serving cell information) and adapts trans-

mission strategy based on current conditions. Table 1 presents the adaptive thresholds and corresponding strategies for each network type.

3.5 Dual-Layer Buffer System

A two-tier buffering architecture ensures zero message loss during network outages. The primary buffer uses a Python collections.deque structure in Random Access Memory (RAM) with a capacity of 10,000 messages. When RAM buffer utilization exceeds 80%, messages automatically overflow to the secondary SQLite database, providing persistent storage for extended outages. Upon network recovery, messages are transmitted in First-In-First-Out (FIFO) order with configurable batch sizes (default: 50 messages) to prevent network congestion during buffer flush operations.

4 Experimental Methodology

4.1 Test Environments

Comprehensive testing was conducted across different network types:

- 5G-NSA: Deutsche Telekom network with EN-DC (LTE Band 3 + NR Band n78)
- 4G LTE: Deutsche Telekom Band 3 (1800 MHz)
- Campus Wi-Fi: Eduroam
- Private Campus 5G-SA: BTU campus network Band n78

4.2 Test Protocols

We implemented a livestock sensor data simulator which generated health telemetry including temperature (38-40°C), rumen pH (6.2-7.0), and activity levels (0-100) for 72 to 200 virtual cattle. Each simulated cow transmitted data every 60 seconds via MQTT QoS 1. Three test categories were performed: (1) 60-minute baseline tests measuring latency via Internet Control Message Protocol (ICMP) ping to Google DNS (8.8.8.8), sampled every 10 seconds which yielded 360 samples per test; (2) throughput stress tests with 200 cows (double baseline load), and (3) buffer tests using a three-phase protocol: normal operation, simulated outage via remote broker termination, and recovery monitoring.

5 Results

Network	Latency (ms)	Std Dev	P95 (ms)	RSRP (dBm)	Msg Loss
5G-NSA (n78/B3)	38.5	±5.7	44.9	-86.7	0%
4G LTE (B3)	34.7	±4.9	40.0	-85.2	0%
Wi-Fi (2.4 GHz)	24.0	±1.2	26.4	-65*	0%

** Wi-Fi signal measured as RSSI (Received Signal Strength Indicator)*

Table 2: Network Performance (60-minute baseline tests)

5.1 Network Latency Comparison

Table 2 summarizes network performance across 60-minute baseline tests on Deutsche Telekom's commercial network. All networks achieved sub-50ms 95th percentile latency, meeting requirements for real-time livestock monitoring where sub-minute alert delivery is acceptable. Notably, 4G LTE (34.7ms mean) outperformed 5G-NSA (38.5ms mean), which could be attributed to dual-connectivity coordination overhead in NSA mode where the device must coordinate between LTE anchor and NR secondary cells. Wi-Fi achieved the lowest latency (24.0ms) with exceptional stability (±1.2ms standard deviation) due to local network proximity.

5.2 Signal Range Validation

Our testing across multiple environments validated gateway operation across 47 dB signal range. Table 3 summarizes results from BTU campus tests, demonstrating successful operation from good signal conditions in the 5G-SA laboratory (-66 dBm RSRP) to marginal commercial cellular coverage outdoors (-112 dBm RSRP).

Network	RSRP (dBm)	Duration	Status
5G-SA (n78)	-66	~10 min	Operational
Wi-Fi (Eduroam)	-75*	~7 min	Operational
5G-NSA (n78/B3)	-110	~3 min	Operational
4G LTE (B3)	-112	~9 min	Operational

Table 3: Signal Range Validation

5.3 Throughput Scalability

Baseline tests achieved approximately 62 messages per minute (3,720 messages per hour) across all network types. Stress testing with 200 simulated cattle (double the baseline load) showed the gateway achieved sustained throughput of 121.4 messages per minute over a 5-minute test period with zero message loss. Mean latency during stress testing was 36.2ms (±4.0ms), demonstrating no degradation under increased load.

Signal Quality	RSRP (dBm)	Throughput (msg/hr)
Excellent (BTU 5G)*	-66	1,167
Good (Wi-Fi)	-65	3,714
Fair (4G LTE)	-85	3,714
Poor (5G-NSA)	-87	3,720
Marginal (5G-NSA)*	-110	1,101

**Short duration tests, extended validation pending*

Table 4: Performance Across Signal Strength Range

5.4 Buffer Resilience

Buffer resilience was validated using the dual-broker architecture with a 60-second simulated cloud outage. During the 60-second outage, the gateway continued receiving telemetry via the local broker and buffered 51 messages in RAM. Upon broker restoration, the gateway automatically

reconnected and flushed all buffered messages within 2 seconds, achieving 100% message delivery with no data loss. The SQLite overflow mechanism provides additional resilience for extended outages exceeding RAM buffer capacity.

6 Conclusion

This paper presented an adaptive 5G gateway for smart livestock monitoring. Key findings from our tests include obtaining sub-50ms P95 latency across all network types (5G-NSA: 44.9 ms, 4G LTE: 40.0 ms, WiFi: 26.4 ms), with 4G LTE outperforming 5G-NSA possibly due to dual-connectivity overhead. We observed 47 dB operational signal range (-65 to -112 dBm), 121.4 msg/min throughput with 200 cattle and zero message loss, and 100% message recovery within 2 seconds. The results demonstrate that existing 4G/Wi-Fi infrastructure effectively supports livestock IoT, while multi-network adaptability ensures reliable operation across diverse conditions. Future work will include extended 5G-SA and Wi-Fi validation and field trials with real data from bolus sensors.

Acknowledgments

This work was funded by the BMFTR funded project “Innovationscampus Elektronik und Mikrosensorik” (grant 16ME0420K).

References

[1] Neethirajan, S.: The role of sensors, big data and machine learning in modern animal farming. Sensing and Bio-Sensing Research, vol. 29, 2020, Article ID: 100367.

[2] Qazi, S.; Khawaja, A.S.; Farooq, Q.U.: IoT-equipped and AI-enabled next generation smart agriculture: A critical review, current challenges and future trends. IEEE Access, vol. 10, 2022, pp. 21219-21235.

[3] Tang, W.; Zhang, A.; Xiao, J.; Liang, W.; Yu, B.: A survey on the 5G network and its impact on agriculture: Challenges and opportunities. Computers and Electronics in Agriculture, vol. 180, 2021, p. 105895.

[4] Mekki, K.; Bajic, E.; Chaxel, F.; Meyer, F.: A comparative study of LPWAN technologies for large-scale IoT deployment. ICT Express, vol. 5, no. 1, 2019, pp. 1-7.

[5] Talavera, J.M.; et. Al.: Review of IoT applications in agro-industrial and environmental fields. Computers and Electronics in Agriculture, vol. 142, 2017, pp. 283-297.

[6] Wolfert, S.; Ge, L.; Verdouw, C.; Bogaardt, M.-J.: Big data in smart farming - a review. Agricultural Systems, vol. 153, 2017, pp. 69-80.

[7] Liu, Y.; Ma, X.; Shu, L.; Hancke, G.P.; Abu-Mahfouz, A.M.: From Industry 4.0 to Agriculture 4.0: Current status, enabling technologies, and research challenges. IEEE Transactions on Industrial Informatics, vol. 17, no. 6, 2021, pp. 4322-4334.

[8] Meng, Z.; Guo, Y.; Sun, C.; Wang, B.; Sherry, J.; Liu, H. H.; Xu, M.: Achieving consistent low latency for wireless real-time communications with the shortest control loop. Proc. ACM SIGCOMM, New York: ACM, 2022.

[9] Alnaas, M.; Alhodairy, O.: Comparison of 5G Networks Non-Standalone Architecture (NSA) and Standalone Architecture (SA). International Journal of Computer Science Engineering Techniques 8 (2024).

[10] Jafri, S. T. A.; Ahmed, I.; Ali, S.: Queue-Buffer Optimization Based on Aggressive Random Early Detection in Massive NB-IoT MANET for 5G Applications. Electronics 11 (2022) H. 18, 2955.

[11] Raspberry Pi Ltd.: Raspberry Pi 5 Model B Datasheet. Release 1. Cambridge: Raspberry Pi Foundation, September 2023. Online: https://datasheets.raspberrypi.com/rpi5/raspberry-pi-5-product-brief.pdf

[12] Quectel Wireless Solutions: RM520N-GL 5G Module Hardware Design Manual. Version 2.1. Shanghai: Quectel Wireless Solutions, July 2022. Online: https://www.quectel.com/product/5g-rm520n-gl.

DOI 10.5162/iCCC2026/P40

MooNet: Harnessing a Low-Power Mesh Backbone for Livestock Localisation and Monitoring

Navya Sneha Chatharajupalli, Randolf Rotta, and Jörg Nolte,

Chair of Distributed Systems/Operating Systems, BTU Cottbus-Senftenberg
{chathara, rottaran, joerg.nolte}@b-tu.de

Abstract

Continuous monitoring of livestock health enables farmers to quickly detect and react to threats. For example, bolus sensors in the rumen stomach are widely used for cows in combination with location tracking information. Most bolus solutions use LoRa radios to transmit the sensor data and use the received signal strength for localization. However, the number of LoRa gateways needed for area coverage and localization together with their energy demands makes outdoor operation prohibitively costly. This paper examines the combination of low-power LoRa receivers with a low-power mesh network for long-term operation. In particular, it analyses the impact of different communication protocol designs on the power consumption. By synchronizing the activity schedules of sensors and mesh nodes, the proposed design achieves an average power consumption as low as 6mW, while remaining scalable to 1000 sensors. The respective significant reduction in solar panel and battery size makes it possible to economically operate such receivers outdoors.

1 Introduction

Bolus sensors [1] inside the rumen stomach of cows enable health monitoring in order to identify and address critical health conditions early. Table 1 summarizes main features of a selection of bolus sensors available in the market. For communication, all recent systems use radios based on the Semtech LoRa modulation in the 868/915 MHz band. In order to receive the sensor data, most of the systems use full LoRaWAN gateways. Their large antennas help to cover wide areas and their software-defined receiver can decode LoRa signals without knowing the modulation parameters in advance. However, these are expensive and consume a lot of power. In contrast, the SmaXtec gateways use inexpensive LoRa radios. They compensate the reduced range by deploying more gateways and using a custom protocol with fixed modulation parameters.

Although LoRa is designed for communication beyond 10km, the attenuation caused by the cow body reduces the range to at most 500m. In practice, higher bitrates with ranges around 50–100m are used in order to reduce collisions and limit the bolus transmitter's energy consumption. The limited range together with the gateways' dependency on wall-power confines bolus-based health monitoring mostly to indoor operation, as installing and maintaining outdoor power cables on large farmland is expensive.

Company	SmaXtec	Moonsyst	Cattle Scan	HKT
Dimension	132x35mm	100x32mm	130x32mm	130x32mm
Range	50–100m	400m	300m	500m
Battery life	5 years	6 years	6 years	5 years
Radio	LoRa modulation @ 868/915 MHz			
Receiver	LoRa radio	LoRaWAN gateway		

Table 1 Example Bolus Sensors in the market along with the maximum coverage and communication technology.

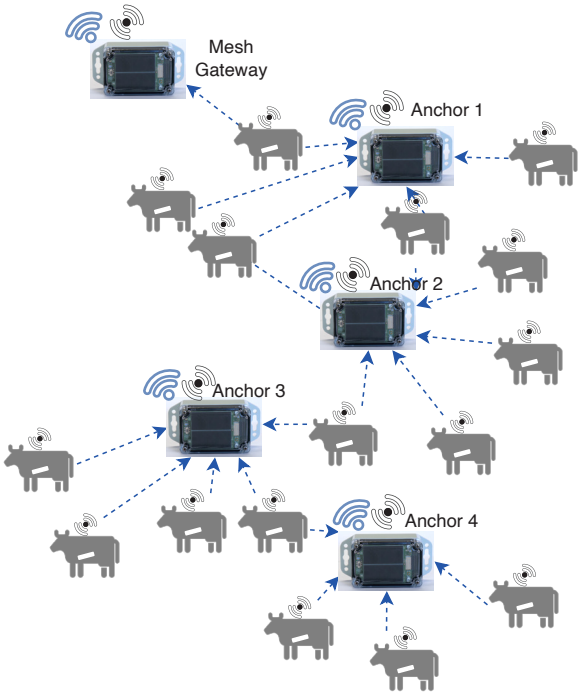


Figure 1 Architecture of a solar-power mesh network with boli inside the cows roaming in the farm in between the mesh nodes (Anchors)

Therefore, this paper examines the feasibility of battery- and solar-powered outdoor operation. Figure 1 depicts the LoRa receivers as anchors of the location tracking. Messages from a bolus are received by all nearby anchors and are then reported to the internet gateway together with the received signal strength (RSSI) and anchor position. To cover extensive farmland, long range communication to the gateway is needed, which can be accomplished with the LoRa radio of the anchors. However, forwarding all bolus messages via multiple anchors will lead to significant medium contention and will violate the airtime fairness regulations. To overcome this, we propose to use a high-

DOI 10.5162/iCCC2026/P40

throughput, low-power mesh network based on IEEE 802.11 radios on 2.4GHz as communication backbone.

In the past, the most power-efficient solution for such mesh networks were Raspberry Pi Zero W single-board computers using the 802.11s or BATMAN mesh protocol. Their power consumption ranges from 400 to 2000mW and, for example, a solar panel of 20W peak power, i.e. a size around 40x40cm, would be a reasonable choice. However, installing and maintaining many of such large solar panels on the farmland is also expensive.

In order to significantly reduce the size of the solar panels, we combine low-power multi-radio microprocessors such as the Espressif ESP32-C6, a custom multi-hop mesh routing protocol [8,9], and clock synchronized radio duty cycling. Although the receive and transmit power consumption of the 802.11 radio is similar to single-board computers, significant energy savings are attainable by exploiting the very fast sleep and wake-up transitions. To this end, we present different communication protocol designs and analyze their impact on the average power consumption.

The analysis shows, that sleep scheduling based on synchronization across all anchors and boli allows to reduce the average power consumption of the anchor nodes down to 6mW. This makes it possible to reduce the solar panel to 200mW peak power or less, which translates into a size of just 4x3cm. In consequence, anchor node and solar panel can be combined into a weather-resistant housing that is easy to install and maintain even on extensive farmlands.

The next section summarizes related work. Section 3 describes the proposed hardware architecture and derives key power consumption parameters. Section 4 introduces the protocol designs and analyses their average power consumption based on performance modelling.

2 Related Work

Early research publications focused on developing wireless sensor networks for cattle health monitoring [2]. Considering the frequency selection, penetration depth and bandwidth earlier designs focused on wearable collars for the cattle. For instance, [3] used a combination of omni-directional and directional ceramic patch antennas for RF-based non-invasive sensing. Although these collars achieved a good penetration depth at 315MHz, the relationship between signal properties and cow health was inconclusive. Later, wireless communication was integrated into bolus sensors that are placed directly inside the cow's stomach [4]. These bolus sensors provide information such as body temperature, heart rate, pH level, rumination, and stress [1]. Finally, [5,6,7] provide models for the attenuation of ultra-high frequency waves through the cow body.

With the increasing availability of inexpensive long-range technologies such as Semtech's LoRa modulation and the LoRaWAN protocol, these were also adopted for bolus sensors. As visible in Table 1, many products now use LoRaWAN in the 868/915 MHz band, although the signal attenuation is worse than on the 315/433 MHz bands.

3 A Low-Power Mesh Backbone for Bolus Anchors

Various transceivers with support for the LoRa modulation exist and, out of these, modules with an integrated low-power microcontroller can simplify the development of the anchor nodes. One example is the RAK wireless RAK3172 module [12], which is based on the STM32WLE5CC microcontroller and has variants for all subGHz bands.

Many mesh technologies are available for the communication backbone between anchors and internet gateway. Based on the observation that the reasonable maximum distance between bolus and anchors is around 100m, the maximum distance between anchors would not be much higher. The objective of the backbone is to maximize the throughput at that distance: Higher throughput finishes transmissions faster and, hence, allows to switch back to low-power sleep earlier. IEEE 802.11, also known as Wi-Fi, achieves this by providing a wide range of bitrates and supporting large frames, e.g. 1.5kB and more, for reduced overhead.

In order to bring 802.11-based mesh networking and low-power microprocessors together, we implemented and improved the BATMAN-V protocol for the ESP32 modules family [8,9]. For power efficiency, the variants ESP32-C3, C6, and the new C5 are most interesting.

Table 2 summarizes the power consumption of different operating modes according to the manufacturer's specifications. These numbers already include the power of an idle processor. Own measurements with the Nordic Power Profiler Kit confirmed the order of magnitude. The ESP32-C6 [11] consumes around 257mW (78 mA at 3.3V) with active receiver and up to 1.3W during transmissions at maximum transmit power. Note, that typical transmissions take just 0.2–1ms even for large frames because of the high bitrates and, hence, are quite energy efficient. The RAK3172 module [12] consumes around 17mW (5.22mA at 3.3V) with active receiver.

From this we can estimate the average power consumption of an anchor node. For example, listening continuously on both radios will cost around 274mW, which is already better than the 1W baseline from a Raspberry Pi Zero mesh node. Duty cycling the 802.11 radio to 1s per minute while listening continuously on LoRa would reduce the consumption down to 4.3+17≈21mW.

Mode	Receive	Transmit	Processor	Sleep
Wi-Fi	257mW	1250mW	46–125mW	23–594µW
LoRa	17mW	287mW	0.7–12mW	6µW

Table 2 Anchor power consumption with different radio operation modes based on the data sheets.

4 Low-Power Protocol Designs between Boli and Anchors

This section discusses and analyzes the protocol designs with focus on the communication between anchors and bolus sensors. For the mesh backbone we assume that a 1/60 duty cycle with 4mW average consumption is sufficient.

The first subsection describes the design assumptions and three protocol designs. Then, the second subsection compares their power consumption in relation to the number of bolus sensors or farm animals. The last subsection discusses the impact on the power consumption of the bolus.

4.1 Protocol Designs

The protocol design must consider the following criteria:

- Each bolus broadcasts a status messages every 15 minutes containing information such as bolus identifier, temperature, rumen activity indicators, pH level, and battery state. Allowable airtime for the LoRa frequency band chosen.
- If the bolus sensors indicate a critical condition, an alert message is broadcasted every 1 minute. The probability for such condition is very low, for example 5%.
- For location tracking, every anchor listens for every bolus independent of past locations, because the animal can have moved far since the last message.
- All received messages are forwarded as soon as possible to the internet gateway. We do not focus on data compression by aggregating measurements over time because this could delay the delivery of important information.

Protocol 1 is the simplest design: The anchors listen all the time for LoRa messages from any bolus. The bolus broadcasts its messages with a random offset in order to minimize collisions. Thus, the average consumption of the anchor is 4.3mW for the mesh backbone plus 17mW for LoRa.

Protocol 2 is inspired by the BLE periodic advertisements [14]. Every bolus starts at a randomized time but sends its messages in fixed periodic intervals. Initially, the anchors listen in order to learn the individual bolus schedules and broadcast this information to all other anchors. After all sensors are discovered, the anchors activate the LoRa receiver only shortly for potential transmissions from each bolus.

This reduces the LoRa listening time. However, each anchor still has to listen for each bolus's alert message once per minute although these are very unlikely. Because of necessary guard times around these listening intervals, the actual duty cycle of the receiver reaches 100% quickly with growing number of boli. In addition, the activity interval of two boli can overlap due to relative clock drift and, hence, result in continued message collisions.

Protocol 3 aims to reduce the listening time for unlikely alarm messages. This is achieved by sharing a few short time slots for alarms across all boli as shown in Figure 2. Collisions in these time slots are reduced through LoRa Channel Activity Detection [13] with a small random backoff. Furthermore, an exclusive periodic time slot is assigned to each bolus for its normal status messages.

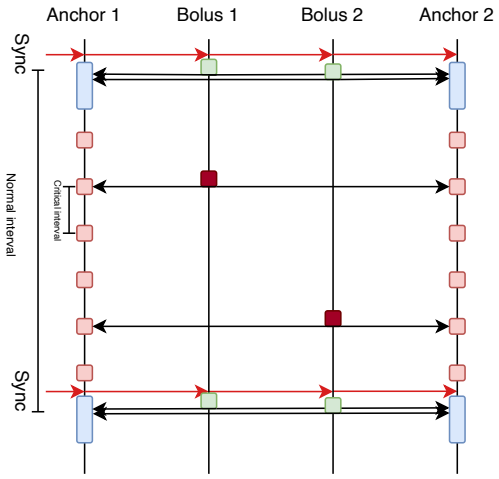


Figure 2 Global schedule for exclusive normal status messages every 15 minutes and shared short time slots for alarm messages every minute.

In order to establish the schedule, all anchors are first synchronized through the mesh network. Then, periodic beacon broadcasts from the anchors allow the boli to synchronize too. The shared alert time slots are also be used by the boli to request their exclusive time slot. Although much more complex, the main advantage is that a small number of alarm time slots every minute is sufficient.

4.2 Protocols power consumption vs number of cows in a farm

We created a power consumption model for the anchor's LoRa radio based on the consumption values from Table 2. The power consumption of LoRa transmissions is estimated based on airtime calculations using [10] and the average number of transmissions per hour. For this, we assumed a payload size of 24bytes, 250KHz bandwidth, coding rate 4/5, and spreading factor 9.

The LoRa listening intervals are also based on the airtime calculations and include an additional guard time before and after in order to compensate for clock drift between bolus and anchor. The model does not account for overlapping listening intervals because this is not relevant for worst-case estimates.

Figure 3 shows the average power consumption depending on the number of cows from 1 to 1000. Most farms or pastures in Germany have less than 500 cows [15]. For protocol 1, the typical power consumption is 21.3mW irrespective of the number of cows. Protocol 2 has a lower power consumption of 4.4mW for a single cow. However, the consumption increases and around 130 cows it reaches protocol 1. Because of the shared alarm time slots, the consumption grows much slower for protocol 3. It reaches 6mW for 1000 cows in our model.

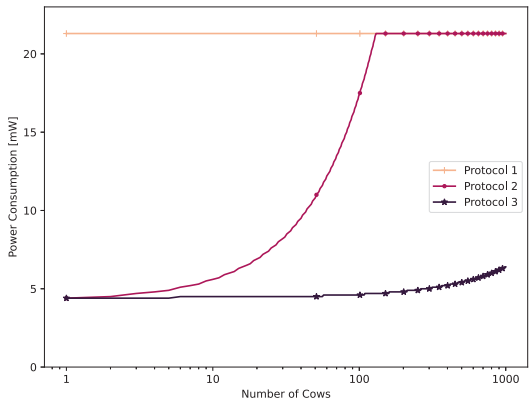


Figure 3 Anchor power consumption for the three protocols and different pasture size. For readability, the horizontal axis is on a logarithmic scale.

4.3 Bolus Energy Budget

The bolus is inserted into the cow's rumen and stays there forever. Thus, its battery has to provide sufficient energy for the desired 5–6 years of lifetime, see Table 1. In the following we analyze only protocol 3 because it has the most activities for the bolus.

The energy consumption of sending the normal messages every 15 minutes is 0.787mWh per day based on the airtime calculator. In order to receive the beacon broadcasts from anchors, the bolus has to listen once every 15 minutes including guard times, which costs 0.209mWh per day.

The alert messages could be sent every minute but we assume that not all boli have alerts all the time. We considered pessimistically, that each bolus has 5 days of alerts in a month, which would cost in average 1.9mWh per day. This calculation also includes the listen before talk time with channel activity detection.

Thus, the estimated total energy consumption of the bolus is 2.96mWh per day. For a lifespan of 5 years, this would amount to 5.4Wh just for radio operation excluding sensors and local preprocessing. For reference, a single Li-FeS2 AA battery has a capacity of 4.5Wh (3000mAh at 1.5V) and a typical bolus has sufficient space for three of such batteries, that is 13.5Wh.

5 Conclusions

In this paper, we proposed a communication protocol that utilizes clock synchronization to enable low-power communication between cattle bolus sensors and nearby anchor nodes. Compared to the gateways available on the market, this can reduce the consumption from 1W to 6mW. This enables the application of solar harvesting with a small solar panel, for example with 200mW peak power and 4x3cm size. This makes it economically suitable, to install and operate a sufficient number of anchors for outdoor livestock monitoring and localization tracking.

Acknowledgements

This work was funded by Bundesministerium für Forschung, Technologie und Raumfahrt (BMFTR) in the project Innovationscampus Elektronik und Mikrosensorik Cottbus – iCampμs under project number 16ME0420K.

6 References

[1] A.H. Ipema, *et al.*: Pilot study to monitor body temperature of dairy cows with a rumen bolus. Computers and Electronics in Agriculture, Volume 64, Issue 1, 2008, Pages 49-52,

[2] Mayer, K.; Ellis, K; Taylor, K.: Cattle Health Monitoring using Wireless Sensor Networks. Proceedings of the Communication and Computer Networks Conference (CCN 2004), S. 8-10.

[3] Andonovic, I. *et al.*: Wireless Sensor Networks for Cattle Health Monitoring. In: Davcev, D., Gómez, J.M. (eds) ICT Innovations 2009. ICT Innovations 2009. Springer, Berlin, Heidelberg.

[4] Sharma, B. and Koundal, D.: Cattle health monitoring system using wireless sensor network: a survey from innovation perspective. IET Wirel. Sens. Syst., 2018 8: 143-151.

[5] Nadimi E.: Modelling Wireless Sensor Networks for Monitoring in Biological Processes. Aalborg: Department of Control Engineering, Aalborg University, 2008. 151 p.

[6] Mimouni A.: Design and Performance Analysis of a Miniaturized Low-Profile Antenna for Cattle Health Monitoring Tags. *J. Phys.: Conf. Ser.* 3027 012047

[7] H. Dogan, I. B. Basyigit, M. Yavuz and S. Helhel.: UHF wave attenuation throughout the cow body. *2017 10th International Conference on Electrical and Electronics Engineering (ELECO)*, Bursa, Turkey, 2017, pp. 1046-1049.

[8] R. Rotta, *et al.*: B.A.T.M.A.N. Mesh Networking on ESP32's 802.11. *2024 IEEE 49th Conference on Local Computer Networks (LCN)*, Normandy, France, 2024, pp. 1-7

[9] R. Rotta, *et al.*: Demo: B.A.T.M.A.N. Mesh Routing on Ultra Low-Power IEEE 802.11 Modules. *2024 IEEE 49th Conference on Local Computer Networks (LCN)*, Normandy, France, 2024, pp. 1-4

[10] LoRa: https://www.semtech.com/products/wireless-rf/lora-connect/sx1276

[11] Esp32 datasheet: https://www.espressif.com/sites/default/files/documentation/esp32-c6_datasheet_en.pdf

[12] RAK3172: https://docs.rakwireless.com/product-categories/wisduo/rak3172-module/datasheet/

[13] LBT-CAD: https://www.semtech.com/uploads/technology/LoRa/cad-ensuring-lora-packets.pdf

[14] BLE-periodic-advertisements: https://www.bluetooth.com/blog/periodic-advertising-sync-transfer/

[15] BMEL-Statistik: https://www.bmel-statistik.de/landwirtschaft/tierhaltung/rinderhaltung

DOI 10.5162/iCCC2026/P41

Schnelle Überprüfung miniaturisierter optischer Transceiver mithilfe Neuronaler Netze

Yingbin Xu[1)], Rainer Tutsch[2)], Kai Schmieder[1)], Tobias Schneider[1)], Rene Kirrbach[1)]
1) Fraunhofer Institut für Photonische Mikrosysteme, Dresden, Deutschland
2) Institut für Produktionsmesstechnik, TU Braunschweig, Braunschweig, Deutschland

Kontakt: yingbin.xu@ipms.fraunhofer.de

Kurzfassung

Optische Transceiver können hinsichtlich ihrer Funktion detailliert untersucht und charakterisiert werden. Dafür werden i. d. R. spezielle Messaufbauten benötigt und die Messungen selbst sind zeitaufwändig. In diesem Beitrag wird eine schnelle Detektionsmethode auf Basis von Convolutional Neural Networks (CNNs) zur Bewertung der Übertragungsqualität optischer Transceiver vorgestellt. Erste Ergebnisse an verschiedenen Transceiver-Typen zeigen, dass diese Methode die Positionierung der Linse relativ zur Laserdiode mit mikrometergenauer Präzision vorhersagen kann. Damit lässt sich die Prüfzeit pro Bauteil von etwa 15 Minuten für eine vollständige Bewertung auf konventionellem Wege auf rund 1 Minute senken. Gleichzeitig wird mit dieser neuen, KI-gestützten Prüfung eine bestehende Lücke im Repertoire schneller Testmethoden geschlossen.

1 Einführung

Die im Beitrag erwähnten Transceiver bestehen aus einem Sender für eine erste optische Datenübertragungsstrecke und einem Empfänger für einen zweiten optischen Link. Zum Senden der Daten werden beispielsweise Laserdioden und Fotodioden zum Empfang der optischen Signale eingesetzt. Die Strahlformung geschieht durch Linsen, die auf den jeweiligen Einsatzfall angepasst sind. Selbst geringe Positionierungenauigkeiten der optischen Komponenten können den Querschnitt des emittierten Strahls erheblich verändern und die Systemleistung beeinträchtigen und zu einem sprunghaften Anstieg der Datenübertragungsfehler führen. Herkömmliche optische Messverfahren mit Lineartischen liefern präzise Ergebnisse; ihre langen Messzyklen machen sie jedoch für die großflächige Online-Inspektion in Produktionslinien ungeeignet. Convolutional Neural Networks (CNNs) haben in den vergangenen Jahren verbreiteten Eingang in Bilderkennung und industrieller Inspektion gefunden.

Dieser Beitrag beschäftigt sich mit einer automatisierten Inspektionslösung für den industriellen Einsatz. Mithilfe von Strahlbildern wird die relative Lageabweichung der Linse zum Substrat in Echtzeit bestimmt. Das gefundene Verfahren arbeitet berührungslos. Im vorliegenden Beitrag wird ein industriell einsetzbarer End-to-End-Workflow beschrieben. Positionierabweichungen bei optischen Transceivern können auf Basis von Kamerabildern und CNNs regressiv vorhergesagt werden. Als Datenbasis dienen optische Simulationen sowie experimentell aufgezeichnete Strahlprofile. Auf Modellseite kommt ein Regressionsrahmenwerk zur Anwendung, das eine multi-skalige Merkmalsextraktion mit geometrischen Vorinformationen kombiniert.

2 Methodik

2.1 Datenbasis

Das Training neuronaler Netze erfordert eine große Anzahl von Strahlprofilbildern als Trainingsdaten. Für das Generieren dieser Daten wird die Simulations-Software: „ZEMAX OpticStudio“ eingesetzt. Das Koordinatensystem wird so gelegt, dass der Ursprung mit der Position der Laserdiode im Transceiver-Modell übereinstimmt.

Für die Simulation des optischen Systems wird das so genannte Raytracing [01] genutzt. **Bild 1** zeigt beispielhaft eine solche Simulation. Für eine angenommene Lageabweichung zwischen 0 und max. 100 µm in x, y und z wurden circa 30.000 Strahlprofilbilder erzeugt. Drei davon sind exemplarisch in **Bild 2** gezeigt.

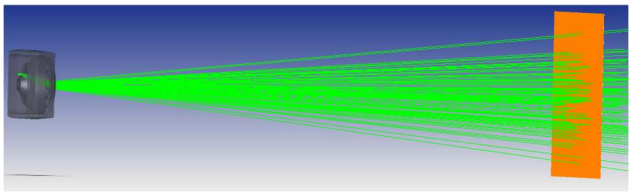


Bild 1 Ansicht einer Raytracing-Simulation mit Austrittsfenster (links) und Detektorfläche (rechts)

Der simulierte Strahl durchläuft eine definierte Detektorebene. Diese entspricht der Kamera im Experiment.

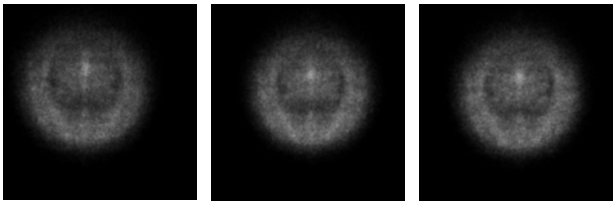


Bild 2 Mit ZEMAX simulierte Strahlprofile (Beispiel)

DOI 10.5162/iCCC2026/P41

2.2 Bildvorverarbeitung

2.2.1 Filtern

Die Intensität der Strahlprofile *I* ist bei den untersuchten optischen Transceivern im Zentrum am höchsten und nimmt typischerweise nach außen hin ab. Zur Rauschunterdrückung und Verbesserung der Konturen wurde der in Gleichung (1) angeführte Gauß-Filter eingesetzt:

$$I_{fil}(x) = \frac{1}{W(x)} \sum_{\zeta \in N(x)} G_{\sigma_s}(\| \zeta - x \|) G_{\sigma_r}(|I(\zeta) - I(x)|) I(\zeta) \quad (1)[02]$$

Dieser verwendet einerseits einen räumlichen Gauß'schen Filterkern $G_{\sigma_s}(\| \zeta - x \|)$, dessen Wirkung vom Abstand des Pixels zum Zielpixel x abhängt und andererseits einen weiteren Gauß'schen-Filterkern $G_{\sigma_r}(|I(\zeta) - I(x)|)$, der die Differenz der Pixel-Intensitäten gewichtet. *W(x)* gibt die Normierung an. ζ steht für das Nachbarpixel.
Auf diese Weise bleiben die Kanten erhalten, während zufälliges Rauschen reduziert wird.

2.2.2 Bild-Binarisierung

Im nächsten Schritt muss eine Schwellwert-Segmentierung gemäß Gleichung (2) erfolgen. Ein mit einem geeigneten Schwellenwert T erzeugtes Binärbild kann den Strahlquerschnitt deutlich vom Hintergrund unterscheiden:

$$I_{binary}(x,y) = \begin{cases} 1, & I(x,y) \geq T \\ 0, & I(x,y) < T \end{cases} \quad (2)$$

Das resultierende Binärbild idealer Transceiver zeigt einen annähernd kreisförmigen Bereich. Im Gegensatz zu den simulierten Bildern weisen reale Profile häufig eine ungleichmäßige Intensitätsverteilung auf. Dies kann zu Konturbrüchen oder unscharfen Kanten führen. Um diese Problematik zu beheben, wird ein iteratives Schwellwert-Verfahren auf Basis der Bild-Binarisierung eingesetzt [06].

$$Q = \frac{A}{U} \quad (3)$$

$$C = \frac{1}{Q} \quad (4)$$

In den o. g. Gleichungen steht *Q* für Qualität, *A* für die Fläche, *U* für den Umfang und *C* für die Kosten. Gleichung (3) ist zu entnehmen, dass ein Profil mit einem hohem Qualitätswert ein größeres Flächen-Umfangs-Verhältnis aufweist und damit näher an einer Kreisform liegt. Je höher die Qualität des Strahlprofils, desto geringer erscheinen die Kosten. So der Schwellwert *T* in Gl. (2) wird so gewählt, dass die Kosten C minimal sind. **Bild 3** zeigt das Effekt-Diagramm und die Profilqualität bei gewählten Schwellwerten *T* von 15, 40 und 60.

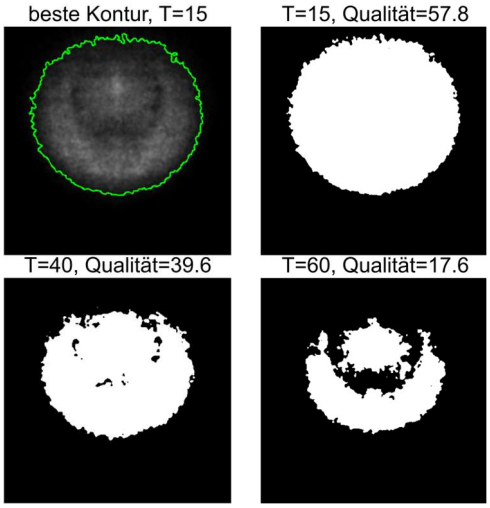


Bild 3 Abhängigkeit der Qualität des Strahlprofils von der Wahl des Schwellwertes *T* - 15 (oben rechts), 40 (unten links) und 60 (unten rechts)

Die vorgenommene Implementierung verwendet einen iterativen Schwellwert-Algorithmus, um den Schwellwert-Bereich anzupassen. Dabei werden alle zur Auswahl stehenden Schwellwerte hinsichtlich der erzielbaren Profil-Qualität getestet [07].

2.2.3 Kanten-Erkennung

Bei der Kanten-Erkennung wird der so genannte Canny-Algorithmus [03] angewendet, um die Konturen weiter zu optimieren. Bei diesem Algorithmus handelt es sich um ein Standardverfahren der digitalen Bildverarbeitung von Graustufenbildern. Die Berechnung der Intensitätsänderung im Bild erfolgt mittels Sobel-Operatoren, die die Gradienten in x- und y-Richtung bestimmen. In den Gleichungen (5) und (6) steht *G* für den Gradienten, *S* für der Sobel-Operator und *EB* für das Eingangsbild.

$$G_x = S_x * EB = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} * EB \quad (5)$$

$$G_y = S_y * EB = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix} * EB \quad (6)$$

Anschließend wird der Bereich mit den markantesten Änderungen ermittelt, der die Kante des Strahlprofils darstellt. Im verwendeten Programm-Code sind zwei Schwellwertparameter enthalten, um die Intensität der Strahlkante zu bestimmen zu können. Wenn der Betrag des Gradienten größer als der obere Schwellwert ist, wird dies als Kante betrachtet. Liegt sie niedriger als der untere Schwellwert, wird keine Kante erkannt [04].

2.2.4 Bild-Dilatation

Bei der Kanten-Erkennung kann es zu Unterbrechungen oder Fehlstellen (beispielsweise in Form von Löchern) im Profil kommen. Um diese zu schließen und um die Strahlkontur insgesamt homogener darzustellen zu können, wird Methode der Dilation eingesetzt.

DOI 10.5162/iCCC2026/P41

Die Dilatation kann als die so genannte Minkowski-Summation des Strukturelements SE_B auf dem Zielgebiet $_A$ betrachtet werden, d. h. als Vereinigung aller Translationen von SE_B mit SE_A ($SE_A \oplus SE_B$). Dadurch wird die Grenze von SE_A nach außen hin verschoben. Auch kleinere Fehlstellen (Löcher) können so gefüllt werden.
Bild 4 zeigt das ursprüngliche Binarisierungs-Ergebnis eines simulierten Strahlprofils. Im Strahlprofil auf der linken Seite sind noch Lücken erkennbar. Für die mittlere Abbildung wurde ein kreuzförmiges 9×9-Strukturelement zur Ausdehnung verwendet, wobei bevorzugt die horizontale und vertikale Richtung erweitert werden. Die kleinen Lücken werden dadurch im Wesentlichen gefüllt. Die diagonalen Linien des Profils weisen jedoch noch Kanten auf. Die rechte Abbildung verwendet ein kreisförmiges Strukturelement gleicher Größe zur Ausdehnung. Die isotrope Ausdehnung glättet die Kanten und schließt gleichzeitig die verbleibenden Lücken vollständig, wodurch der Fülleffekt deutlicher wird.

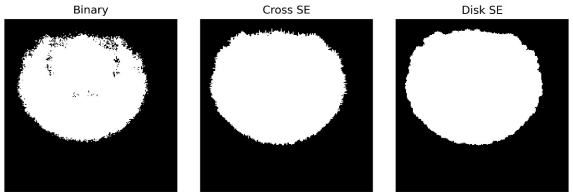


Bild 4 Verdeutlichung des genutzten Dilatationsprinzips: Lücken nach Binarisierung (links), Anwendung eines kreuzförmigen Strukturelementes (Mitte) und eines kreisförmigen Strukturelementes (rechts)

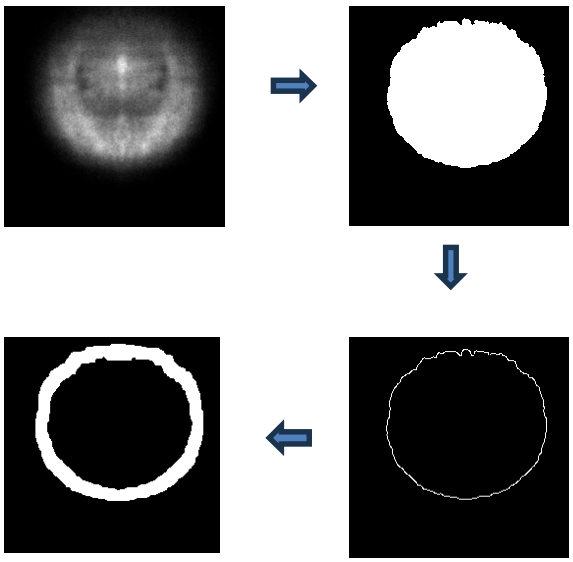


Bild 5 Verdeutlichung der Bildverarbeitungsschritte: gefiltertes Bild (links oben), binäres Bild (rechts oben), Bild mit Kantenerkennung (rechts unten) und dilatiertes Bild (links unten)

In der Bildverarbeitung muss oftmals eine Erosion nach der Dilatation durchführt werden, um eine Vergrößerung des Zielbereichs zu verhindern. Dieser Weg wurde hier nicht beschritten, da der Schwerpunkt in diesem Beitrag auf dem Erkennen des Strahlprofils und der Merkmals-Extraktion liegt. Für die Strahlprofilerkennung ist es besser, das dilatierte Bild direkt zu verwenden. [05]
In der gefilterten Darstellung von **Bild 5** ist zu erkennen, dass die Helligkeit der Kontur im Originalbild von unten nach oben allmählich abnimmt. Dadurch verringert sich der Pixelgradient, und am oberen Rand werden die Grenzen unscharf, sodass keine klare Kantenlinie mehr erkennbar ist. Diesem Effekt wird bei der Kanten-Erkennung Rechnung getragen, und im Dilatationsprozess wird der Kontrastverlauf gezielt verstärkt, indem helle Pixelbereiche weiter ausgedehnt und dunkle Randzonen betont werden, um den Übergang zwischen intensiven und abgeschwächten Strahlungsbereichen präziser herauszuarbeiten.

2.3 Merkmalsextraktion

Nach erfolgter Bild-Vorverarbeitung werden im nächsten Schritt Strahlprofil-Merkmale für das Training von Regressions- und neuronalen Netzwerkmodellen ermittelt. Aus dem Strahl-Profil extrahiert man dazu die Merkmale, die zur Charakterisierung von Form, Position und Intensitätsverteilung genutzt werden können. **Bild 6** zeigt die Merkmals-Gewinnung.

2.3.1 Hough-Kreis-Erkennung

Da die eingesetzten Transceiver einen kreisförmigen bzw. elliptischen Strahlquerschnitt aufweisen, kann die so genannte Hough-Transformation eingesetzt werden. Diese basiert auf der Kreisgleichung: [08].

$$(x-a)^2 + (y-b)^2 = r^2 \tag{7}$$

Dabei sind a, b die Koordinaten des Kreismittel-Punkts und r der Radius. In der Praxis werden die Randpunkte des Strahls zunächst mit dem Canny-Algorithmus identifiziert. Für jeden Randpunkt werden mögliche Mittelpunkte berechnet, die im Parameter-Raum Kreisbögen bilden. Der Schnittpunkt der überlagerten Bögen entspricht dem tatsächlichen Mittelpunkt.

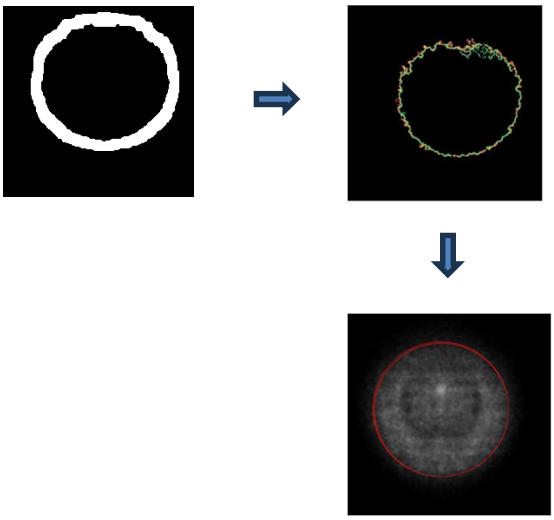


Bild 6 Merkmalsextraktion anhand der Kontur (oben rechts) sowie durch Hough-Kreise (unten rechts)

Um Fehlinterpretationen bei der Vorhersage zu minimieren, werden die drei wahrscheinlichsten Kreise ausgegeben, sodass die realen Profile möglichst genau approximiert werden können [09]

2.3.2 Kontur-Mittelpunkt und Flächen-Berechnung

Nach der Bestimmung des Strahlprofils müssen der Konturmittelpunkt und die Fläche erfasst werden. Der Konturmittelpunkt drückt im gegebenen Anwendungsfall die relative Position der Linse zur Laserdiode aus. Die Fläche repräsentiert den Strahlquerschnitt. Die Quantifizierung dieser Merkmale erfolgt anhand des Konturmittelpunkts und der Profilfläche. Die dazugehörige Berechnungsformel lautet wie folgt [10]:

$$M_{pq} = \sum x^p \, y^q I(x,y) \tag{8}$$

$$x_c = \frac{M_{10}}{M_{00}}, y_c = \frac{M_{01}}{M_{00}} \tag{9}$$

In dieser Gleichung steht M_{pq} für Bildmoment, p ist die Ordnung des Moments in x-Richtung, q die Ordnung in y-Richtung und die Indizes: c für Konturmittelpunkt. Da die kreisförmige Kontur aus einem gefilterten Binärbild entsteht, kann ihr Helligkeitswert I nur die Werte 0 oder 1 annehmen. Daher entspricht der Mittelpunkt der Kontur ihrem Schwerpunkt. In diesem Fall entspricht M_{00} der Summe der Intensitäten und (x_c, y_c) dem geometrischen Schwerpunkt der Kontur, der die relative Lage der Linse zur Laserdiode direkt beschreibt.

3 Innovative Convolutional-Neural-Network

Ungleichmäßige Intensitätsverteilung sowie die teilweisegeringen Kanten-Details im Strahlprofilbild erfordern die Verwendung mehrerer Faltungskernel-Größen damit beide Kriterien im Strahlprofilbild simultan verarbeitet werden können. Aus diesem Grund wird die Verwendung eines innovativen CNN-Modells mit 3-Zweigen vorgeschlagen. Für Zweig 1 kommt ein (3×3-Kernel) zum Herausfinden lokaler Intensitätsmaxima zur Anwendung. Mit diesem erfolgt die Vorhersage der *x*-Koordinate.
Für Zweig 2 ist ein (5×5-Kernel) vorgesehen. Dieser dient der Erfassung der Strahlprofil-Form sowie der Vorhersage der *y*-Koordinate.
Schließlich wird ein (7×7-Kernel) für die Verzweigung 3 eingesetzt. Dieser ist darauf ausgerichtet, die globale Intensitäts-Verteilung zu analysieren und die zusätzlichen Informationen für *x* und *y* zusammenzuführen.

3.1.1 Korrelations-Verbesserung

Zur Verbesserung der Korrelation wird nach der Vorverarbeitung zunächst eine 1×1-Faltung auf das Eingabebild angewendet, deren Ergebnis mittels Sigmoid-Funktion in den Bereich [0,1] transformiert wird und als Gewicht dient. Die gewichteten Merkmale entstehen durch elementweise Multiplikation.

3.1.2 Ausgabeschichten

Bei beiden Zweigen 1 und 2 handelt es sich um vollverbundene Schichten, die die *x*- bzw. *y*-Koordinate ausgeben. Schließlich wird F_3 mit den Merkmalen aus den Zweigen 1 und 2 fusioniert. Dabei konzentriert sich Zweig 1 auf lokale Details (*x*), Zweig 2 auf die Konturform (*y*), während Zweig 3 globale Zusatzinformationen liefert.

3.1.3 Verlustfunktion

Um die Unabhängigkeit der Zweige sicherzustellen, werden separate Verlustfunktionen definiert. Zusätzlich wird ein intensitätsgewichteter Gradienten-Verlust verwendet, um die genaue Lokalisierung der *x*- und *y*-Grenzen zu unterstützen. Dieser berechnet die Abweichung zwischen dem vorhergesagten Gradienten und dem wahren Gradienten. Damit kann dem Modell mitgeteilt werden, wo Anpassungen vorgenommen werden müssen. Die Umsetzung erfolgte in Anlehnung an die in [03] vorgestellt Methode.
Der normalisierte Pixelintensitätswert verstärkt den Einfluss der helleren Bereiche im Strahlprofil. Um eine Überverstärkung zu vermeiden, wird der Maximalwert mit der min-Funktion auf einen Wert zwischen 1 und 2 begrenzt.

3.1.4 CNN kombiniert mit Hough-Transformation

Zur weiteren Verbesserung der Robustheit werden die Ergebnisse der Hough-Kreis-Erkennung aus der Merkmalsextraktion als zusätzliche Eingabe integriert. Der erkannte Konturmittelpunkt (x_c, y_c) wird vor der verbundenen Schicht mit den CNN-Merkmalen verknüpft. Dadurch wird das Modell in die Lage versetzt, geometrische Vorab-Informationen mit den gelernten Bildmerkmalen zu kombinieren und die Vorhersagegenauigkeit für die x- und y-Koordinaten zu steigern.

3.2 Modellarchitektur

Das CNN-Modell gemäß **Bild 7** enthält drei parallele Faltungszweige, eine Korrelationsverbesserung und eine Hough-Merkmalsfusion. Die Eingabeebene umfasst den PNG-Bildeingang und einen Hough-Merkmalseingang. Das PNG-Bild wird mithilfe einer dreistufigen Architektur verarbeitet.
Aufgabe der ersten Faltungsschicht ist das Extrahieren lokaler Merkmale mithilfe verschiedener Faltungskerne. Die Batch-Normalisierungsschicht soll die Merkmalsverteilung vereinheitlichen und den Trainingsprozess beschleunigen. Mit Hilfe der Max-Pooling-Schicht wird u. a. die Translationsinvarianz verbessert.
Die zweite Faltungsschicht dient der Erfassung lokaler Formen und Konturen. Anliegen der Korrelations-Verbesserung ist die Verstärkung wesentlicher Bild-Merkmale und die Unterdrückung nebensächlicher Informationen. Die nachfolgende Schicht führt ein 2×2-Pooling zur weiteren Komprimierung durch und reduziert das Bild anschließend zu einem eindimensionalen Vektor *F1* bis *F3*.

DOI 10.5162/iCCC2026/P41

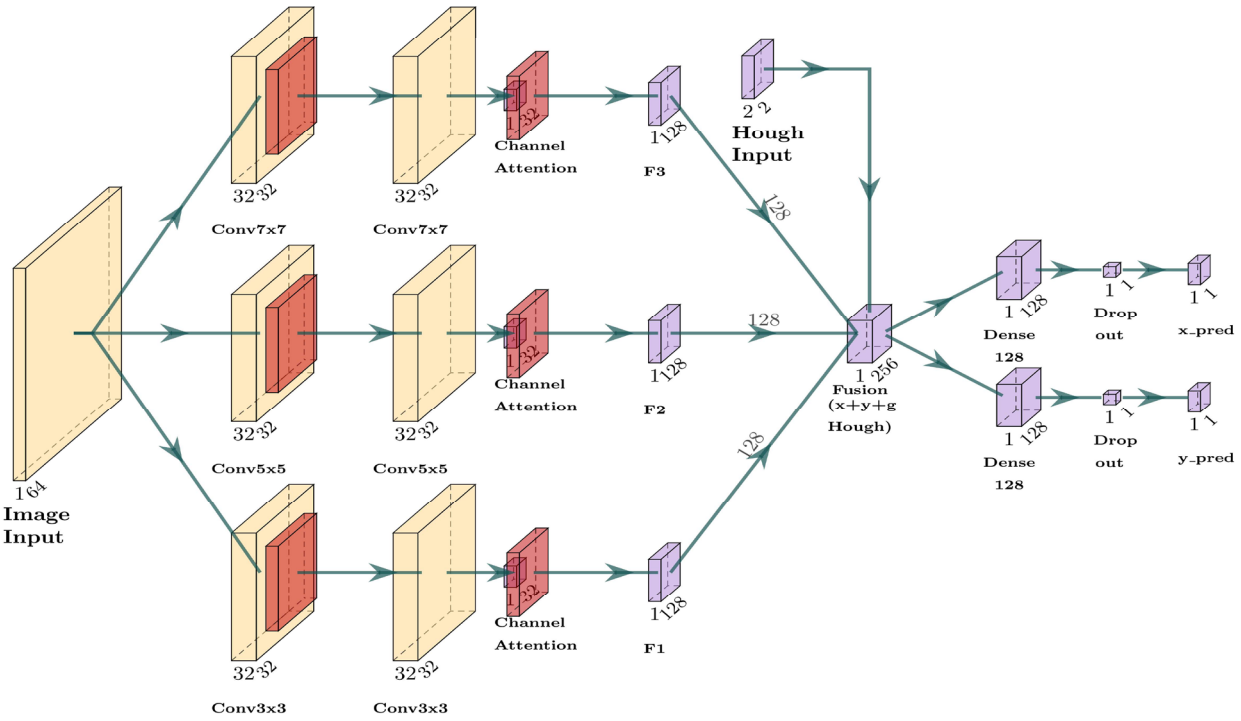


Bild 7 Gewählte Struktur des 3-zweigigen CNNs

Die reduzierten Merkmale *F1*, *F2* und *F3* der drei Zweige werden über die Verkettungs-Ebene mit dem Hough-Eingang fusioniert, um einen umfassenden Merkmalsvektor zu erhalten. In diesen werden mehrskalige Bildmerkmale und geometrische Vorhersagen integriert. Die fusionierten Merkmale werden schließlich von zwei parallelen Ausgabeköpfen verarbeitet, um die *x* bzw. *y*-Koordinaten vorherzusagen.

4 Test und Validierung

4.1 Ergebnisse der Vorhersage

Bild 8 veranschaulicht den Zusammenhang zwischen dem vom Modell vorhergesagten und dem Zielwert für Δx. Die gestrichelte Linie repräsentiert den idealen Vorhersagewert. Die Verteilung der Datenpunkte um diese Linie zeigt die gute Anpassung des Modells an den Zielwert. Eine nahezu identische Aussage trifft auch auf die prognostizierten Werte für Δy zu.

Eine beispielhafte Darstellung für konkrete Versatzwerte eines Transceiver-Samples ist in **Bild 9** gezeigt. Aus der GUI lässt sich im ausgewählten Fall auf eine Lageabweichung Abweichung von 6 µm in x und 8 µm in y-Richtung schließen.

Die durchgeführten Experimente zeigen, dass das entwickelte CNN-Modell die Vorhersage der ΔX- und ΔY-Koordinaten mit hoher Genauigkeit und Robustheit leisten kann. Die kompakte Verteilung der Datenpunkte entlang der idealen Linie bestätigt die effektive Anpassung des Modells an die simulierten Montage-Abweichungen.

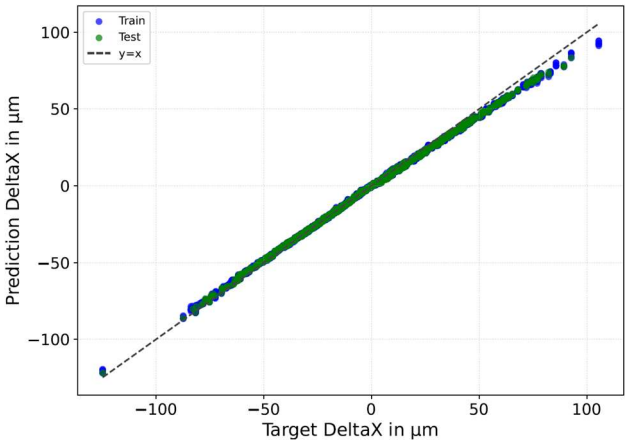


Bild 8 Sollantwort für die Δx-Komponente (CNN mit 3 Verzweigungen)

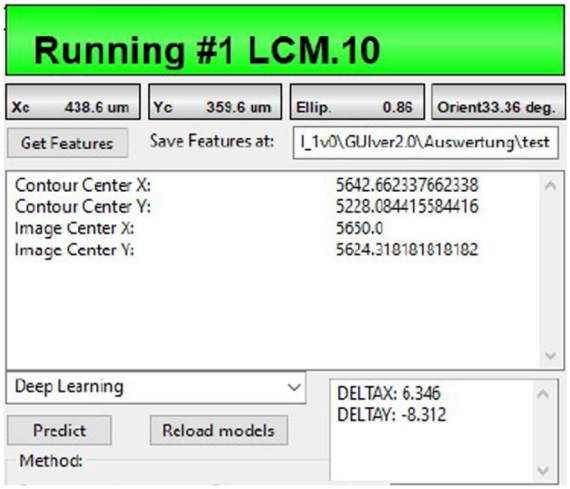


Bild 9 Beispiel für die ermittelten Versatzwerte eines Transceivers

4.2 Diskussion

Die Ergebnisse zeigen, dass die Kombination aus mehrskaliger Faltung, Korrelations-Verbesserung und Hough-Merkmals-Fusion die Vorhersageleistung des Modells signifikant steigert und im Vergleich zu traditionellen Methoden die Robustheit und die Anpassungsfähigkeit des Testverfahrens an unterschiedliche Strahlprofil-Formen erhöht. Die Einbeziehung geometrischer Vorinformationen spielt eine entscheidende Rolle bei der Lokalisierung der Koordinaten.

Nichtsdestotrotz weist auch das hier vorgestellte Modell Einschränkungen auf. Beispielsweise hängt seine Leistungsfähigkeit stark von der Qualität der simulierten Daten ab.

5 Fazit und Ausblick

Dieser Beitrag schlägt eine innovative CNN-Architektur vor, die durch die Fusion von Bild-Merkmalen mit geometrischen Informationen nach Hough eine hochpräzise Vorhersage von Punktkoordinaten ermöglicht. Zukünftige Arbeiten konzentrieren sich auf die folgenden zwei Bereiche:

- Die Entwicklung einer automatisierten Pipeline zur signifikanten Vergrößerung des Trainingsdatensatzes ermöglicht die Generierung zahlreicher, leicht variierter Trainingsbeispiele aus wenigen realen Bildern.
- Methoden zur Übertragung simulativ trainierter Modelle auf reale Daten und zur Anpassung an deren spezifische Eigenschaften werden untersucht (z. B. Transferlernen und Domänenanpassung). Dadurch wird der Übergang von der Simulation zur Realität beschleunigt und die schnelle Modellierung unterschiedlicher Transceiver-Typen unterstützt。

6 Literatur

[01] J. R. Mahan, The Monte Carlo Ray-Trace Method in Radiation Heat Transfer and Applied Optics, ASME/Wiley, 2019, pp. 1–28.

[02] Carlo Tomasi und Roberto Manduchi. “Bilateral filtering for gray and color images”. In: Sixth international conference on computer vision (IEE Cat.No.98CH36271). IEEE. 1998,S. 839 – 846 S.2

[03] John Canny. “A computational approach to edge detection”. In: IEEE Transactions on pattern analysis and machine intelligence 6 (1986), S. 679 – 698 S.6-8

[04] I. Sobel and G. Feldman, “A 3×3 isotropic gradient operator for image processing,” Stanford Artificial Intelligence Project, 1968

[05] R. C. Gonzalez und R. E.Woods. Digital Image Processing. 3. Aufl. Upper Saddle River, NJ: Prentice Hall, 2008 S.343-360

[06] Sonka, M., Hlavac, V., & Boyle, R. (2014). Image Processing, Analysis, and Machine Vision (4th ed.). Cengage Learning S.248

[07] Jähne, B. (2005). Digital Image Processing (6th ed.). Springer S.427

[08] Hough, P. V. C. (1962). “Method and means for recognizing complex patterns.” S.1

[09] L. Bachmann, Entwicklung eines Software Tools zur computergestützten Auswertung von Transceiver Profilen auf Basis Neuronaler Netze für die Prozessqualifizierung, Diplomarbeit, Institut für Halbleiter- und Mikrosystemtechnik, Technische Universität Dresden, Apr. 26, 2021. S.46

[10] R. C. Gonzalez und R. E. Woods. Digital Image Processing. 3. Aufl. Upper Saddle River, NJ: Prentice Hall, 2008 S.688

DOI 10.5162/iCCC2026/P42

Formation and role of the interface region in ultrathin metal oxide layers for resistive gas sensors

Carlos Morales; Rudi Tschammer; Dominic Guttmann; Karsten Henkel; Jan Ingo Flege* (Brandenburgische Technische Universität Cottbus-Senftenberg, FG Angewandte Physik und Halbleiterspektroskopie, 03046 Cottbus, Germany)
Carlos Alvarado; Christian Wenger (IHP – Leibniz-Institut für innovative Mikroelektronik, Im Technologiepark 25, 15236 Frankfurt (Oder), Germany)
* flege@b-tu.de

Abstract

In miniaturized gas resistive-sensor systems, understanding sensor performance and determining the optimal thickness of the active material requires an in-depth characterization of material deposition. In particular, the early stages of growth are mediated by interface formation and film/substrate interaction, which influence the mode of growth and the physico-chemical properties of the active deposit, likely affecting the sensor response for ultrathin layers. We present the use of *in-situ* X-ray photoelectron spectroscopy and *operando* spectroscopic ellipsometry for characterizing thin (<20 nm) atomic layer-deposited layers for use in next-generation miniaturized sensor devices. By targeting cerium and tin oxide layers, we demonstrate how these techniques can provide insights into their material composition, thickness, and optical properties during interface formation and the transition to bulk-like properties.

1 Introduction

The hydrogen economy proposes replacing current fossil fuels with H_2 produced from water splitting using electricity from renewable energy sources. However, the H_2/air mixture is explosive at ambient pressure for concentrations between 4% and 75%. Therefore, to establish H_2 as a reliable and cost-effective fuel, we require safety elements such as H_2 detectors that can be easily integrated into mass devices.

Resistive sensors made from metal oxides have garnered significant attention due to their numerous advantages, including ease of thin-film nanopatterning, compatibility with integrated circuits, excellent performance with high sensitivity, and rapid response and recovery times. However, meeting market expectations remains challenging. This class of sensors requires either high-temperature heating for operation (>250 °C), which increases power consumption and compromises long-term stability, or the use of expensive noble metals, which raises manufacturing costs. As a result, current research is concentrating on two key areas: (i) creating new sensing materials that are highly responsive at low temperatures without expensive noble metal catalysts, and (ii) finding deposition methods that can create conformal, ultrathin layers on nanopatterned substrates with high effective surface area that can be easily integrated into CMOS technology.

In this context, atomic layer deposition (ALD) is a prominent technique in microelectronics, offering exceptional control over film thickness and conformality, including complex nanostructured surfaces, as well as precise growth of mixed compounds or heterostructures on demand. This is achieved by alternating pulses of organometallic precursors and oxidants, which can only react with the available surface sites in a self-limiting way. Moreover, the relatively low temperatures used for ALD growth, which are compatible with CMOS fabrication processes, yield defect-rich and nonstoichiometric thin films. Far from being a disadvantage, these intrinsic defects may induce modified material properties compared to physical vapor deposited (PVD) counterparts and can even be beneficial in terms of increasing reactivity and thus sensor performance.

ALD processes are particularly suitable for growing ultrathin deposits, i.e., those with a thickness of a few tens of nanometers. However, such thin layers are critically influenced by the interaction with the substrate and the formation of complex interfaces, which likely affect the sensor response for ultrathin layers (< 5nm). Therefore, in miniaturized resistive-sensor systems, understanding sensor performance and establishing the minimum optimized thickness of the active material requires an in-depth characterization of material deposition, particularly at the early stages of growth. We present a combination of *in-situ* X-ray photoelectron spectroscopy (XPS) and *operando* spectroscopic ellipsometry for characterizing ultrathin (<20 nm) atomic layer-deposited layers of cerium and tin oxide (CeO_x and SnO_2, respectively). We demonstrate how these techniques can provide insights into composition, thickness, and optical properties of the deposited materials during interface formation.

2 Experimental methods

Our homemade ALD reactor consists of a stainless steel ultra-high vacuum compatible chamber with a base pressure of 10^{-8} mbar, which is directly attached to the preparation chamber of the XPS system. The sample is positioned at the focal point of the chamber, below the entrance of the reactive gases, and its position can be precisely controlled through a 3-axis linear manipulator with a θ rotator. The sample stage is capable of heating the sample to a maximum temperature of 1000 °C using a silicon nitride ceramic heater (343-HEATER SIN-8X10, Allectra). The temperature is monitored using a K-type thermocouple located near the sample and cross-checked against a temperature-dependent optical model for a Si(111) reference sample, as measured by ellipsometry, between 100 and 200 °C.

DOI 10.5162/iCCC2026/P42

After the ALD growth, and once the pressure is below 10^{-7} mbar in the ALD reactor, the sample is transferred to the XPS analysis chamber. More details can be found elsewhere [1].

The cerium oxide (CeO_x, 1.5<x<2.0) layers were deposited using two different organometallic precursors by thermal ALD. First, by using the commercial precursor $Ce(thd)_4$ (Epivalance) in combination with ozone (~7%, O_3/O_2 mixture) generated with an ozone generator (OXP-30, Oxidation Technologies) and with a substrate temperature of 250 ºC [2]. Second, by using $Ce(dpdmg)_3$ - elaborated and provided by the group of Prof. Devi at Leibniz Institute for Solid State and Materials Research (Dresden, Germany) [3] - in combination with either O_3/O_2, O_2, or H_2O at a substrate temperature of 160 ºC. The tin oxide (SnO_2) was grown using the commercial precursor TDMASn (STREM) and O_3/O_2, at a substrate temperature of 160 ºC. In all cases, the ALD reactor was operated in flow-type mode using N_2 as a carrier gas.

The growth of our ALD-active films has been systematically studied using *operando* spectroscopic ellipsometry (SER 801 UV-VIS, SENTECH) and in *situ* by X-ray photoelectron spectroscopy (XPS, Omicron EA 125) [1]. The inelastic peak shape analysis of XPS survey spectra was performed using QUASES software. The combination of these techniques comprises chemical and structural information of the ALD key actors (precursors, substrate, and film) during the early stages of growth, identifying key factors that can affect the sensing performance of ultrathin thin films such as changes in the growth per cycle (GPC), complex interface formation, and gradients in the oxidation state of reducible metal oxides.

3 ALD-based cerium oxide

3.1 Hetero- and homodeposition regimes

The facile exchange between Ce^{3+} and Ce^{4+} oxidation states under reducing or oxidizing conditions makes CeO_x a promising candidate material for catalysis and sensor applications. For ALD films, the previously introduced intrinsic chemical and structural defects, along with film/substrate interactions, have been shown to critically affect their initial oxidation state [2] as well as their reducibility and sensing behavior towards H_2 [4].

The deposition of ALD-based CeO_x using $Ce(thd)_4$/ozone on SiO_2 substrates illustrates the complex shift in the reaction mechanism that may happen as the precursor first interacts with the substrate and then with the deposit. This behavior defines two distinct growth stages easily differentiated by *in-situ* XPS characterization. During the initial heterodeposition regime (CeO_x directly grown on the substrate), the growth is dominated by non-ALD side reactions on the substrate, which, in the present case, result in greater GPC values and carbon residues, and prevent complete oxidation of the cerium oxide. In contrast, the steady homodeposition state stage (CeO_x grown on CeO_x) is marked by a constant, literature-consistent GPC close to 0.2 Å/cycle,

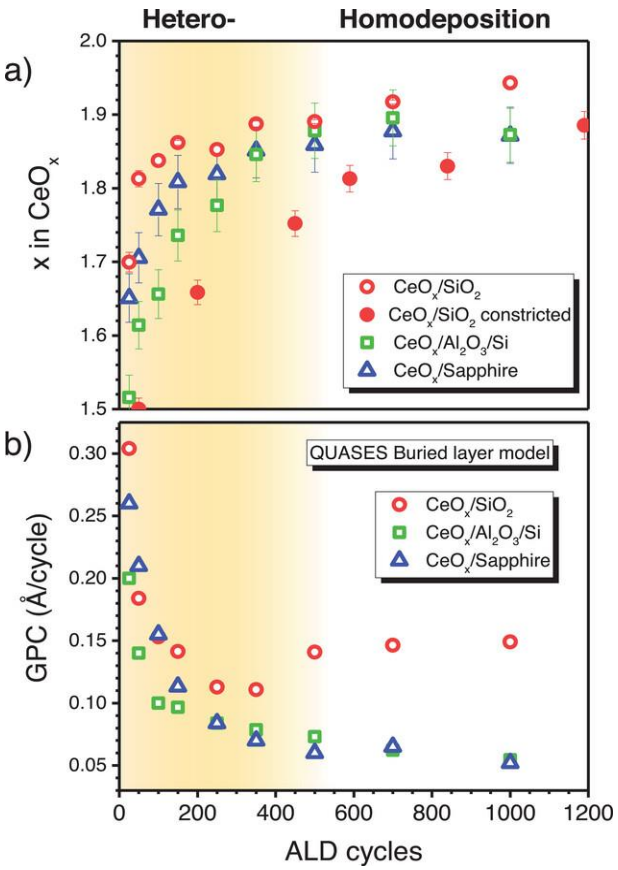


Figure 1. a) Stoichiometry of CeO_x deposits estimated from XPS Ce 3d spectra as a function of the total number of ALD cycles using $Ce(thd)_4/O_3$ (open red circles). The constricted exposure of Ce-precursor is shown by filled red circles. The stoichiometry of CeO_x deposits on ALD-Al_2O_3 on Si and c-sapphire is plotted in open green squares and blue triangles, respectively. b) Growth per cycle (GPC) values obtained from the inelastic peak shape analysis of XPS survey spectra in the Ce 3d region. Reproduced from [2], CC BY license from *Adv. Mater. Interfaces* published by Wiley-VCH GmbH

minimal carbon residues, and a fully oxidized cerium oxide film. These trends are not particular to SiO_2 substrates and can be generalized to other metal oxide supports, such as c-sapphire and ALD-Al_2O_3 (10 nm) grown on Si wafers. **Figure 1** illustrates this shift in the growth mechanism, specifically the (**Fig. 1a**) average amount of Ce^{3+} states (i.e., reduced ceria) and the (**Fig. 1b**) GPC as a function of the number of ALD cycles.

The variation in the GPC can be used to explain the reaction mechanism by applying Puurunen's theoretical model [5], which relates the total amount of metal atoms attached per cycle with the size of the ligands (i.e., steric hindrance effect) and reactivity toward a specific surface (i.e., active sites available). During the homodeposition stage, a ligand exchange reaction is likely taking place: the density of Ce cations attached to the surface after a growth of 0.2 Å/cycle ranges from 0.45 to 0.50 nm^{-2} for $CeO_{1.5}$ and CeO_2, respectively, which are higher than the theoretical value assuming that no ligand release occurs (0.25 nm^{-2}). Moreover, the higher GPC during the heterodeposition regime cannot be

DOI 10.5162/iCCC2026/P42

explained by the steric effect but is likely related to a higher density of initial active sites on the SiO_2 surface, which would promote non-ALD side reactions that increase the presence of carbon residues at the surface, as observed by XPS. Therefore, the initial stages of the growth are significantly affected by the substrate, which induces a change in the reaction mechanism.

More importantly for the chemical behaviour of CeO_x as a sensor active layer, the $Ce^{3+} \rightarrow Ce^{4+}$ transition depends on three factors: (i) the easier reduction of ceria's surface compared to volume; (ii) the interaction with the substrate at the interface region; and (iii) the intrinsic ALD-defects. The first factor is intrinsically related to ceria regardless of the deposition process used. ALD can also enhance this effect by constraining the dose of $Ce(thd)_4$ during the corresponding pulse, thus preventing the complete coalescence of the cerium oxide islands, which now grow in clusters. As shown in **Figure 1a**, this situation leads to higher amounts of Ce^{3+} states compared to the saturated growth. Secondly, the interaction with the substrate might be complex and will depend on the support. For Si-based substrates using PVD methods, it has been reported that the initial formation of silicates results in the stabilization of Ce^{3+} states at the interface [6]. Here, by following the XPS Si 2s region plotted in **Figure 2**, we also observe the formation of such silicates by the appearance of a second component shifted by 1.9 eV to lower binding energies than the SiO_2 line. It is worth noting that, in comparison with PVD processes, the Ce cations in the $Ce(thd)_4$ precursor already exist in the Ce^{4+} oxidation state. Thus, their initial reduction under highly oxidizing conditions (O_3/O_2 at mild temperatures) involves a complex chemical pathway, which is also reflected in the GPC values.

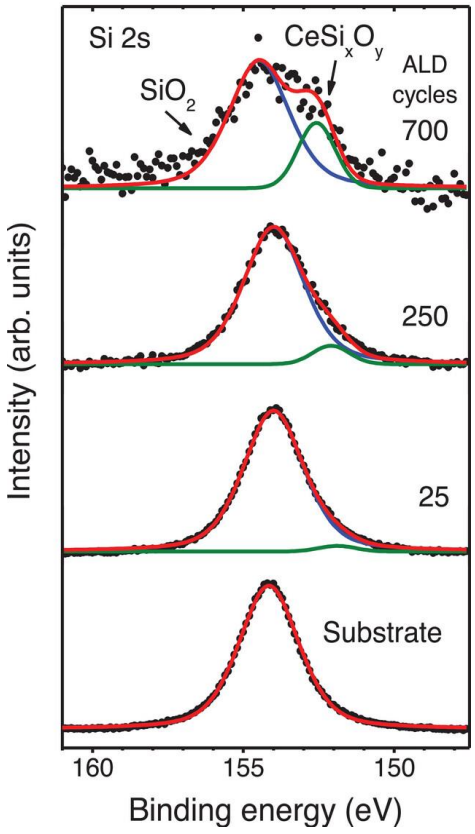


Figure 2. Si 2s XPS spectra as a function of the total number of ALD cycles for the growth of CeO_x on SiO_2 substrates. The red line corresponds to the fit resulting from the Si 2s deconvolution into two components, SiO_2 (in blue) and $CeSi_xO_y$ (green). Reproduced from [2], CC BY license from *Adv. Mater. Interfaces* published by Wiley-VCH GmbH.

As for SiO_2, the use of 10 nm thick ALD-Al_2O_3 (on Si) as a substrate also results in the formation of aluminates at the interface. In this case, however, those Ce^{3+} states are stabilized and cannot be further oxidized when submitted to a post-annealing process at moderate temperatures. These sorts of complementary measurements help establish a bottom-up approach for sensor design by primarily optimizing the active layer. Continuing with this example, the formation of aluminates hinders any gas sensor response when CeO_x films are below 5 nm, thus excluding the use of ALD-Al_2O_3 as the isolating layer between ceria and Si.

3.2 Role of ALD-intrinsic defects in sensing

The intrinsic ALD defects can be observed in two ways. First, by the incomplete oxidation of ceria layers under high oxidizing conditions, as observed for all substrates (the X-ray beam damage is small and can be discarded as a source of reduction). Second, by the lower crystallinity of ALD films. **Figure 3** shows *ex-situ* μ-Raman spectroscopy measurements for a 2.6 nm thick CeO_x film on SiO_2 before (**Fig. 3a**) and after (**Fig. 3b**) a post-annealing process at 250 °C consisting of (1) sample reduction under H_2, (2) subsequently back-oxidizing under an O_2 environment, and (3) final reduction again with H_2. The change in the intensity of the F_{2g} band at ~460 cm^{-1}, in comparison to the Si signal (~300 cm^{-1}), is a fingerprint of ceria crystallization, i.e., the healing of ALD-intrinsic defects, induced by post-annealing. While as-grown ultrathin films have been shown to promote H_2 heterolytic activation at room temperature and in the presence of diluted H_2/O_2 atmospheres (H_2 concentration below 10%), recrystallization translates into lower reducibility, as confirmed by near-ambient pressure XPS measurements [4]. Therefore, ALD-intrinsic defects play a key role in ceria reactivity towards H_2, replicating the expected sensing mechanism at low temperatures.

3.3 Tunable ALD-intrinsic defects

Similar results have been obtained when using the $Ce(dpdmg)_3$ precursor with three different oxidants (O_3/O_2, O_2, and H_2O) and substrates (Si, SiO_2, and ALD-Al_2O_3/Si). While the same general trend regarding the Ce^{3+}/Ce^{4+} ratio is obtained for all growths and substrates as a function of the number of ALD cycles, the amount and variation of Ce^{3+} states heavily depend on the chosen oxidizing species. Similarly, the amount and type of residues (carbon and nitrogen) depend on the oxidizing agent. For instance, and in line with density functional theory (DFT) calculations [3],

DOI 10.5162/iCCC2026/P42

O_2 does not result in full oxidation but rather leaves Ce-O-N bonds, leading to the accumulation of nitrogen in ceria films compared to water or ozone.
Therefore, a careful selection of growth conditions allows for potentially tuning the intrinsic properties of ALD-based films via defect-engineering, including structural (i.e., low crystalline order, amorphous-polycrystalline deposits) and chemical (i.e., incorporation of light elements) defects.

4 ALD-based tin oxide

Lessons learned from the ALD-ceria deposits can now be applied and validated in the study of SnO_x deposits. As determined by *operando* spectroscopic ellipsometry, the GPC is about 0.7 Å/cycle, as expected. In comparison to the ceria growth, no significant change in the GPC has been observed, indicating ideal linear film growth on Si for all cycles (see **Figure 4**). Meanwhile, XPS shows that the film is fully oxidized to SnO_2. Interestingly, notable changes in the estimated Auger parameter during the early stages of growth indicate a shift in the chemical environment when the system transitions from the SnO_2/SiO_2 interface to thicker SnO_2 films. At the same time, preliminary electrical measurements indicate a significant change in resistance and sensing behavior towards H_2 at low temperatures (100 ºC) as a function of film thickness, which could be related

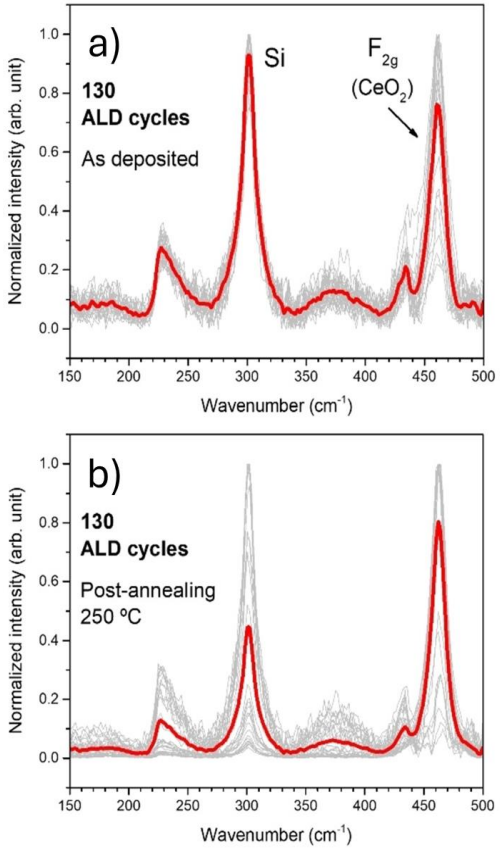


Figure 3. Raman spectra of as-deposited (a) 2.6 nm thick ALD-CeO_x film (i.e. 130 cycles) and after post-annealing at 250 °C (b). The red spectra are the average of several individual spectra (light grey) taken at different positions. Reproduced from [4], CC BY license from *ChemSusChem* published by Wiley-VCH GmbH.

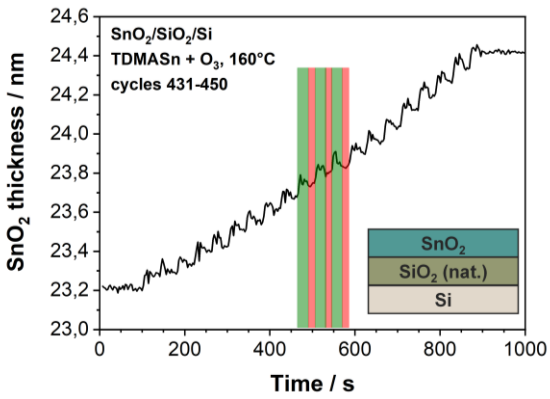


Figure 4. Ellipsometry data of ALD-SnO_2 growth at 160 ºC using TDMASn and O_3/O_2 on Si wafers. Green and red bars indicate the precursor and oxidant doses, respectively. Optical model shown in the right.

to either a change in the amount and type of chemical residues or the interaction with the substrate. Although further experiments are needed, the preliminary results suggest that the SnO_2/SiO_2 interfacial region is a key factor in determining the sensing behavior of ultrathin SnO_2 films.

5 Conclusions

The intrinsic defects of the ALD deposit seem to play a crucial role in the physicochemical properties of ultrathin metal oxide layers. In combination with potentially strong film/substrate interactions, they can be used to tune the material's properties. This opens a new degree of freedom in active layers design while preserving the advantages of thickness and conformability control offered by ALD.

6 Literature

[1] Morales, C.; et al.; Combination of Multiple Operando and In-Situ Characterization Techniques in a Single Cluster System for Atomic Layer Deposition: Unraveling the Early Stages of Growth of Ultrathin Al_2O_3 Films on Metallic Ti Substrates, Inorganics 2023, 11, 477

[2] Morales, C.; et al.: In situ X-ray photoelectron spectroscopy study of atomic layer deposited cerium oxide on SiO_2: substrate influence on the reaction mechanism during the early stages of growth, Adv. Mater. Interfaces 2025, 12, 2400537

[3] Kaur, P.; et al.: Rational development of guanidinate and amidinate based cerium and ytterbium complexes as atomic layer deposition precursors: synthesis, modeling, and application, Chem. Eur. J. 2021, 27, 4913

[4] Morales, C.; et al.: Hydrogen sensing via heterolytic H_2 activation at room temperature by atomic layer deposited ceria, ChemSusChem 2025, 18, e202402342

[5] Puurunen, R. L.: Growth Per Cycle in Atomic Layer Deposition: A Theoretical Model, Chem. Vap. Depos. 2003, 9, 249.

[6] Allahgholi, A; et al.: Oxidation-state analysis of ceria by X-ray photoelectron spectroscopy, ChemPhysChem 2015, 16, 1083

DOI 10.5162/iCCC2026/P43

Cleaning effect on surface activated aluminum to aluminum wafer bonding

Sebastian Schulze[a], Thomas Voß[a], Patrick Krüger[a], Rasuole Lukose[a], Philipp Kulse[a], Matthias Wietstruck[a]
[a]IHP – Leibniz Institut für innovative Mikroelektronik, Im Technologiepark 25, 15230 Frankfurt (Oder), Germany
schulze@ihp-microelectronics.com

Abstract

This work presents a surface activated aluminum to aluminum (Al-Al) wafer bonding process and the effect of different cleaning chemistries used after dry etching on the later bonding quality. Different cleaning procedures based on ammonium fluoride (NH_4F) and a hydroxylamine (NH_2OH) were investigated. The bonding quality was evaluated using C-mode scanning acoustic microscopy (C-SAM) and the results were correlated to the surface properties of the aluminum investigated with atomic force microscopy (AFM), scanning electron microscopy (SEM) and X-ray photoelectron spectroscopy (XPS). We show that the best bonding quality is obtained using a 20 min NH_2OH-based clean. It efficiently removes etch residues without significantly affecting the surface properties of aluminum.

1 Introduction

Al-Al wafer level bonding is a very promising technique for 3D heterogeneous integration. However, a major drawback is the fast oxidation of the Al surface in air. The stable oxide layer acts as a barrier and inhibits the diffusion of Al atoms across the bonding interface [1]. This can be avoided by using a ComBond high vacuum system. Here, the aluminum surface is treated with an Ar plasma, which first removes the oxide and then the wafers are bonded in a high vacuum atmosphere to prevent reoxidation. Still, the bonding result depends heavily on the condition of the aluminum surface. This study investigates the effects of different cleaning methods on the aluminum surface and the subsequent bonding result. The roughness and contamination of the aluminum surface were measured using atomic force microscopy (AFM) and X-ray photoelectron spectroscopy (XPS), respectively. We then correlated these measurements directly with the bonding quality, which was investigated using C-mode scanning acoustic microscopy (C-SAM).

2 Experimental

2.1 Wafer fabrication and cleaning

The top and bottom wafers for Al-Al bonding were prepared as described below. The top wafer contained a non-patterned 1000 nm Al layer (99.5% Al, 0.5% Cu), which was sputtered on a 25nm Ti/ 55nm TiN barrier on top of 600nm SiO_2. The bottom wafer contained the same layer sequence. But the Al layer and the Ti/TiN barrier were additionally patterned using reactive ion etching. The remaining resist was stripped and subsequently the bottom wafer was exposed to different wet cleaning. We investigated NOE cleaning, ACT®935 cleaning of varying lengths, and compared the results with uncleaned wafers. NOE cleaning uses a mixture of ammonium fluoride (NH_4F) and ethylene glycol ($C_2H_6O_2$). It is performed at a temperature of 35 °C for 2:20 min. ACT®935 cleaning uses a mixture of hydroxylamine (NH_2OH) and isopropanol (C_3H_8O). Cleaning is performed at 80°C, and we investigated two different durations: 20 and 40 minutes.

2.2 Surface activated wafer bonding

The surface activated bonding process was carried out in an EVG® ComBond® system using a high vacuum handling cluster with several modules. First, both wafers were transferred to the ComBond® activation module, where a dry etch process removes the oxide and potential contaminants on the Al bonding pads. The pretreatment was performed in an Ar plasma for 3:30 min with a beam current of 100 mA and a voltage of 200 V. The wafers were tilted 15° and rotated with 70 rpm. Without breaking the vacuum to avoid a reoxidation, the wafers were handled to the vacuum alignment module, where the flipped top wafer and the bottom wafer were brought into contact. The wafer stack was then transferred to the bonding chamber where the thermocompression bonding was performed for 1h at 300°C with a force of 60kN.

3 Results

Fig. 1 shows the C-SAM mappings of the bonded wafers using different cleaning methods. In case no cleaning was applied after the resist stripping the C-SAM image reveals several weakly bonded areas (bright areas in fig. 1a). The use of a NOE clean significantly reduces the unbonded regions, though it does not eliminate them entirely (fig. 1b). The best result was obtained with a 20-minute ACT clean (fig. 1c). The few remaining defects in the center of the wafer originate from particles. The two unbonded areas directly at the edge of the wafer are due to the construction of the bottom heater in the bonding chamber. This can be avoided by rotating the wafer by 90 degrees after half of the bonding time. Increasing the ACT®935 process time to 40 minutes does not improve the bonding quality any further (fig. 1c). On the contrary, the C-SAM image reveals larger areas of defects, located in a ring structure close to the wafer edge.
We believe that the large amount of unbonded area for the wafers without additional cleaning is due to residues left

DOI 10.5162/iCCC2026/P43

after the dry etching process, which are visible in the AFM image (fig. 2a). According to the F1s and Al2p XPS spectra shown in fig. 3, these residues are highly oxidized and may contain small amounts of fluorine. During the NOE clean the ammonium fluoride dissolves the native aluminum oxide layer [2] and removes the residues after dry etching process. The AFM measurement (fig. 2b) shows that the average surface roughness (R_a) has increased slightly to 5.1 nm, while the maximum peak-to-valley height (R_z) remains relatively high at 45.8 nm. The SEM image (fig. 2d) of the wafer processed in NOE shows small pits probably created during the cleaning process, leading to a high R_z value and bonding defects. The Al2p XPS spectra in fig. 3b shows a decrease in the intensity of the peak attributed to aluminum oxides. Additionally, a high F1s intensity was detected at a binding energy of 686.4 eV (fig. 3a), which can be assigned to an AlF_3H_2O compound on the surface [3]. After a 20-minute ACT®935 cleaning, the XPS spectra indicate a reduced fluorine concentration and a stronger oxidation of the Al surface. Apparently, this stronger oxidation has no negative impact on the subsequent bonding as the oxide layer is removed during the Ar sputter clean in the high vacuum bonder. The AFM measurement (fig. 2c) shows an average surface roughness (R_a) of 4.6 nm and the lowest R_z value (34.2 nm) compared to the other cleaning variants, which might explain the better bonding quality. A longer ACT clean of 40 min leads to a R_a of 4.1 nm and increases R_z to 51.6 nm (fig. 2d). This might be the reason for the decreased bonding quality.

In summary, we showed that the clean after resist stripping strongly affects the later bonding quality. The best bonding results were obtained after ACT®935 cleaning for 20 minutes, which effectively removes the residues without evident effect on the surface properties of the aluminum.

3.1 Images

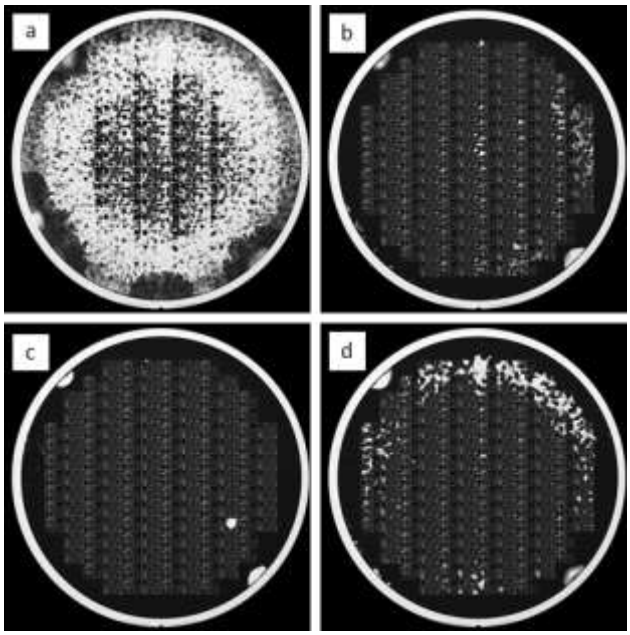


Figure 1 C-SAM images of Al-Al bonded wafers with different cleanings after dry etching: a) no cleaning, b) NOE c) ACT®935 20 min, d) ACT®935 40 min.

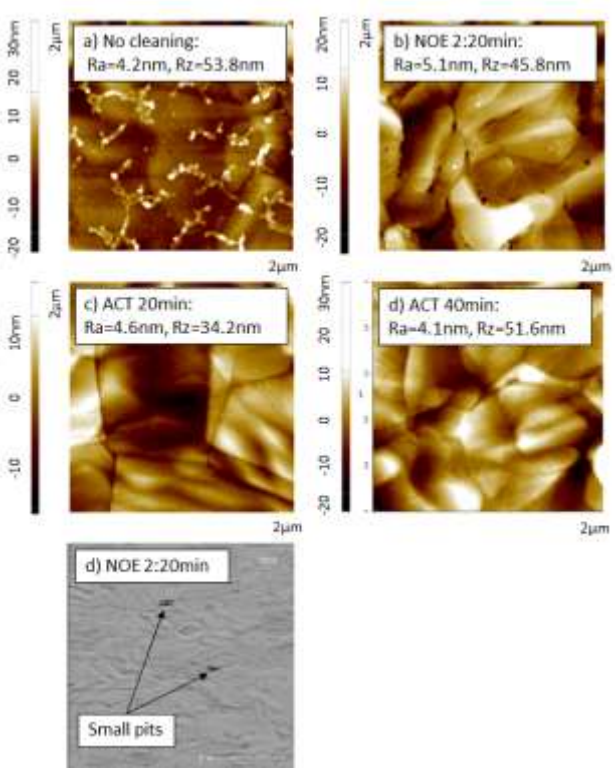


Figure 2 AFM images of wafers with different cleanings after dry etching: a) no cleaning, b) NOE c) ACT®935 20 min, d) ACT®935 40 min, d) SEM image of NOE cleaned wafer.

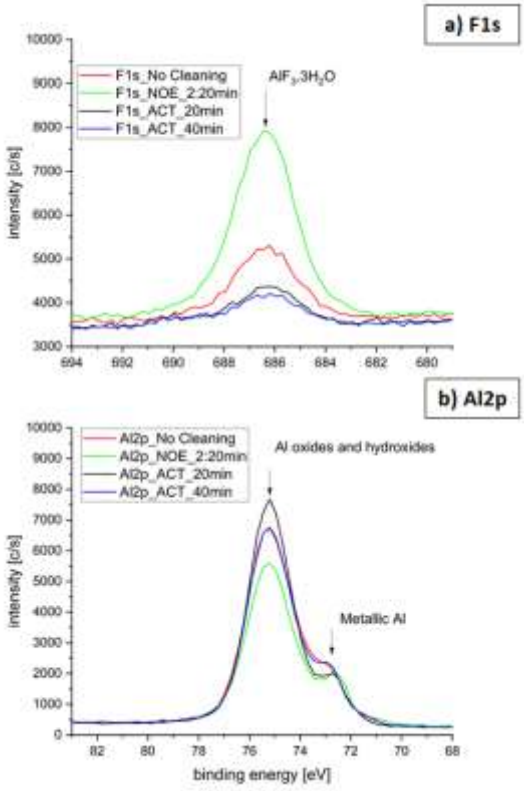


Figure 3 F1s (a) and Al2p (b) XPS spectra of wafers with different cleanings after dry etching.

DOI 10.5162/iCCC2026/P43

4 Literature

[1] A.P. Hinterreiter, B. Rebhan, C. Flötgen, V. Dragoi and K. Hingerl, “Surface pretreated low-temperature aluminum–aluminum wafer bonding”. Microsyst Technol vol. 24, pp. 773–777, 2018

[2] M.K. Carter, R. Small, M. Cernat and B. Hansen, “Effects of Amine Fluoride Cleaning Chemistry on Metallic Aluminum Integrated Circuit Films: I. Experimental Measurements and Chemical Modeling.” Journal of the Electrochemical Society, 150(2), B52, 2003

[3] A. Limcharoen, C. Pakpum, and P. Limsuwan. "An X-ray photoelectron spectroscopy investigation of redeposition from fluorine-based plasma etch on magnetic recording slider head substrate." Procedia Engineering 32, pp. 1043-1049, 2012

DOI 10.5162/iCCC2026/P44

In-situ-Ätzkontrolle für vertikale GaN-basierte Leistungsbauelemente unter Verwendung von Mehrwellenlängen-Interferometrie

In-situ etch control for vertical GaN-based power devices using multi-wavelength interferometry

Ralph-Stephan Unger[1]*, Frank Brunner[1], Enrico Brusaterra[1], Kolja Haberland[2], Sven Peters[3], Olaf Krüger[1]

[1] Ferdinand-Braun-Institut (FBH), Berlin, Germany
[2] LayTec AG, Seesener Str. 10-13, 10709 Berlin, Germany
[3] SENTECH Instruments GmbH, Berlin, Germany

* E-mail: ralph-stephan.unger@fbh-berlin.de

Kurzfassung

Für die Herstellung vertikaler Leistungsschaltbauelemente auf GaN-Basis sind präzise Endpunktdetektionsverfahren für die verwendeten Trockenätzprozesse erforderlich. Eine besondere Herausforderung ist hierbei der geringe Materialkontrast der homoepitaktischen Schichtstapel in den optischen Parametern n und k für verschieden dotierte GaN-Epitaxiestrukturen, die beim Durchätzen nur geringe Unterschiede in den gemessenen Reflektivitäten erwarten lassen. Um das elektronische Verhalten der Driftzone und des p-n-Übergangs in vertikalen GaN-Transistoren zu untersuchen, wurden GaN-Schichten mit unterschiedlichen Dotierkonzentrationen mittels MOVPE (Metalorganic Vapor Phase Epitaxy) gewachsen. Mit Hilfe eines LayTec TRIton in-situ Messsystems mit einer Detektionswellenlänge von 365 nm ist es beim Durchätzen gelungen, Übergänge von unterschiedlich dotierten Bereichen im Epitaxie-Stapel zu erkennen. Die beobachteten Ätztransienten lassen sich qualitativ durch das temperaturabhängige Absorptionsverhalten an der GaN-Bandlücke und den gemessenen Exzitonen-Peak in spektroskopischen Ellipsometermessungen erklären.

Abstract

For the fabrication of vertical GaN-based power switching devices, sensitive endpoint detection methods during dry etching processes are needed. One challenge here is the low material contrast of the homo-epitaxial layer stacks in the optical parameters n and k for differently doped GaN epitaxial structures, which are expected to show only minor differences in the measured reflectance during etching, especially for homo-epitaxial structures on bulk GaN substrates. To investigate the electronic behavior of the drift layers and the p-n junction in vertical GaN transistors GaN layers with different doping concentrations were grown by MOVPE (Metalorganic Vapor Phase Epitaxy). Using a LayTec TRIton in-situ metrology system during patterning of the layer stack by plasma etching with a detection wavelength of 365 nm transitions between differently doped regions in the epitaxial stack could be measured. The observed etching transients can be explained qualitatively by the temperature-dependent absorption behavior at the GaN band gap and by the measured exciton peak in spectroscopic ellipsometer measurements.

1 Motivation

Recently, vertical GaN-based power switching devices on bulk GaN substrates have been developed at FBH for high power density and high switching speed applications [1]. For this purpose, various doped epitaxial layers were grown by MOVPE (Metalorganic Vapor Phase Epitaxy) on 2-inch high-conductivity GaN substrates deploying in-situ monitoring of the reflectance.

The epitaxial structure of a vertical trench MISFET consists in a n^+ doped GaN:Si drain layer, a thick n^- doped GaN:Si drift layer, a p^+ doped GaN:Mg channel layer and a n^+ doped GaN:Si source layer. Further details are described in [2].

For the device fabrication of vertical GaN trench MISFETs, various dry etching steps are needed to etch the recess for the body diode contacts through the top n^+ GaN stopping on the channel layer, to etch the gate trench through the n^+ and p^+ GaN stopping on the drift layer and to etch the mesa edge termination through the three n^+, p^+ and the thick n^- GaN stopping on the n^+ drain layer. All of these etching steps must be done under controlled conditions with in-situ monitoring and precise depth control.

The GaN layers grown on the GaN substrate differ only in terms of their doping concentration and dopant. Due to the small differences in the dopant concentrations only a low contrast in the refractive index can be seen. Additionally,

DOI 10.5162/iCCC2026/P44

the high absorption of GaN on GaN epitaxial structures makes measurement wavelengths above 600 nm useless to detect layers by standard interferometry.

In this work, we demonstrate for the first time the detection of differently doped GaN layers during etching through the entire epitaxial stack, controlled by in-situ interferometry using a LayTec TRIton multi-wavelength in-situ metrology system.

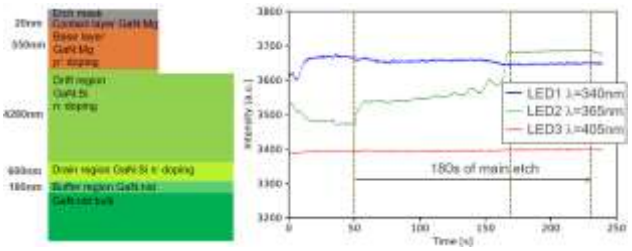


Figure 1 Schematic of the epitaxial test stack (left) used for the gate trench etch process of a vertical GaN trench MISFET. The observed etching transients using a LayTec TRIton multi-wavelength in-situ metrology system are shown on the right.

2 Experiment

A typical GaN epitaxial structure, as shown in Figure 1 on the left, was grown using MOVPE on 2-inch GaN substrates from various suppliers.
To investigate the in-situ measurement capabilities for detecting various doping layers we used a LayTec TRIton in-situ metrology system, equipped with 3 different measurement wavelengths at 340 nm, 365 nm and 405 nm, mounted on top of a SENTECH SI 500 ICP etch chamber. Before starting the etching process, the motorized measuring head was positioned on the unmasked wafer surface inside the chamber. The three wavelengths of the TRIton are focused onto a spot with a diameter of approximately 200 µm, allowing all reflected intensities from the etched wafer surface to be measured simultaneously via an optical path in the chamber.

2.1 Monitoring trench etching

A multi-step chlorine-based ICP process with an etch time of 180 seconds in the main step was used to etch the GaN without surface roughness.
Figure 1 shows the recorded transients of the reflected light from the three LEDs of the TRIton system during etching through the p-n junction.
Due to significant oscillations accompanied by an intensity jump of the etch transient at 365 nm, we determined the boundary between Mg-doped (p~3x10^{19} cm^{-3}) and Si-doped (n~1x10^{16} cm^{-3}) GaN.

2.2 Deep etch at different temperatures

Deeper etching processes through the drift zone are required to etch the mesa edge terminations and form quasi-vertical drain contacts for the transistors. To demonstrate the feasibility of the in-situ measurement system, a complete etch down to the GaN substrate was performed. The result is shown in Figure 2. The reflected intensities at 365 nm wavelength (top curves) show clear transitions depending on the doping concentration. Strongly increasing Fabry-Perot oscillations were observed during transition to the GaN substrate, which can be explained by the decrease in absorption as the layer becomes thinner due to etching. This behavior is temperature-dependent and related to the band gap absorption of GaN [3], which can be observed when etching similar samples at different etching temperatures (see Figure 2).
Neglecting minor changes in etching rate due to temperature, a clear difference in absorption behavior can be observed when similar samples are etched at different etching temperatures (see Figure 3). At an etching temperature of 20 °C, more Fabry-Perot oscillations can be observed, which, compared to etching at 150 °C, is due to a lower extinction coefficient. This is in accordance with the assumption that the GaN band gap absorption shifts to longer wavelengths with increasing temperatures.

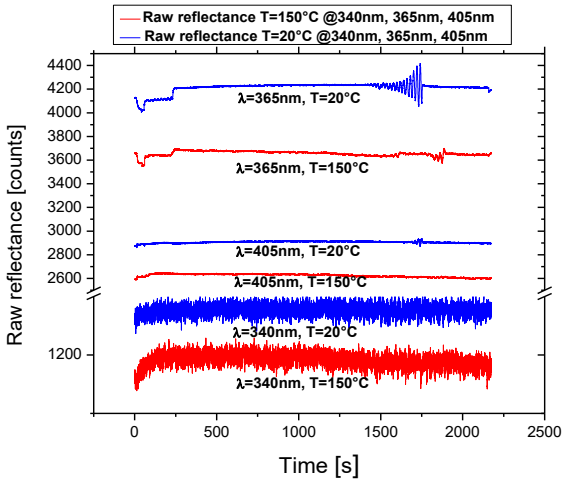


Figure 2 Raw data of etching transients recorded at different etching temperatures on samples from the same epitaxial run.

3 Results and Discussion

To understand the observed features of the measured transients, simulations were performed to predict layer boundaries at a defined etch depth. However, to do so, the knowledge of the optical parameters of the epitaxial layer described with refractive index n and the extinction coefficient k for the used measurement wavelength is the key to simulate the reflectivity vs etch depth.
A spectroscopic SENresearch ellipsometer from SENTECH was used to determine the n and k values based on a describing model. In Figure 3 we compare the refractive index and the extinction coefficient for an etched and the unetched sample.

DOI 10.5162/iCCC2026/P44

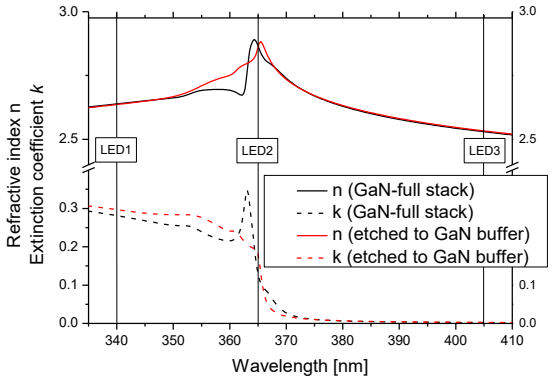


Figure 3 Comparison of optical dispersion based on n and k for an etched and an unetched sample, derived from spectroscopic ellipsometer measurements.

Significant changes in the refractive index and extinction coefficient can be observed very close to the band gap of GaN at 365 nm. This explains qualitatively the observed features in the measured in-situ etching transients only at an emission wavelength of 365 nm (LED2). An exciton peak observed at 363 nm was also found for the unetched sample that is not visible after etching down to the unintentionally doped buffer.

The band gap also depends on temperature and shifts to longer wavelengths as the temperature rises. This behavior was investigated using spectroscopic ellipsometry on a heated sample in which the entire epitaxial stack except for the GaN substrate was etched away. The measurement was performed using the same model, but at different temperatures ranging from 24 °C to 300 °C, as shown in Figure 4.

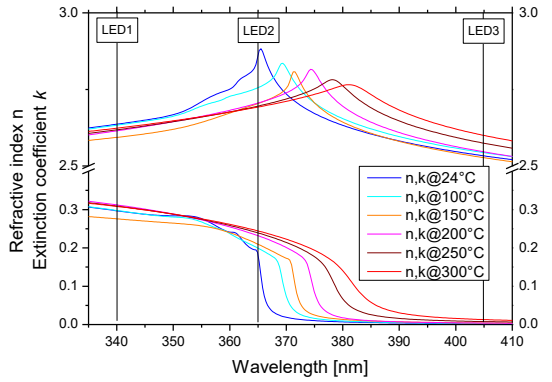


Figure 4 Measured dispersion of refractive index n and the extinction coefficient k based on temperature dependent spectroscopic ellipsometric measurements of an etched sample.

When comparing the n and k values at T = 24 °C and T = 150 °C for a wavelength of 365 nm, we found that the refractive index decreases from 2.85 to 2.70 and the extinction coefficient increases slightly. This is consistent with the observed behavior in Figure 2 and would qualitatively explain the increase in visible oscillation at 20 °C compared to the etching transition at 150 °C. Due to the low material contrast of the doped layers compared to the GaN substrate, it is difficult to separate the n and k values for each layer with a fixed thickness in the model. So, we tried a basic simulation using the transfer matrices model for epi-layers shown in Figure 5 to calculate the expected reflectivity curve in dependence of the etch depth. For the GaN-substrate the n and k values from the etched sample and for the Mg doped GaN epi-layers the n and k values of the unetched sample were used. It has also been reported that absorption increases with the doping concentration of GaN [4]. The n and k values between the top layer and the substrate were obtained by fitting the curve to the measured etch transient at 365 nm wavelength.

The layer boundaries can be predicted very accurately using forward simulation based on 365 nm wavelengths. Comparing the simulation with the data measured at room temperature in Figure 2, the Fabry-Perot oscillations for the p-n junction are more strongly suppressed, but the visible number of Fabry-Perot oscillations corresponds quite well with the two drain layers.

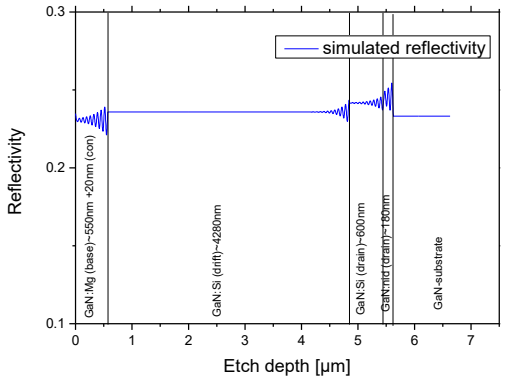


Figure 5 Simulation of reflectivity based on measured n and k values derived from ellipsometric measurements for an etched and an unetched sample.

4 Outlook

A simple simulation based on the optical material parameters n and k for the measurement wavelength of 365 nm used in the transfer matrix model [5] can be applied to qualitatively verify the boundaries between differently doped regions in an epitaxially grown GaN stack on a GaN substrate. In order to understand the change in the visible number of periods in the recorded etch transient curves, the values n and k must be determined more accurately. To do so, we plan to measure the optical parameters for each grown layer separately by etching back the sample used. A more sophisticated approach would be to fit the etch transients using LayTec's SimulyzR software, which can extract the values n and k from the etch data.

5 References

[1] E. Bahat-Treidel *et al.*, "On the Conduction Properties of Vertical GaN n-Channel Trench MISFETs," IEEE Journal of the Electron Devices Society, pp. 1–1, 2021

[2] E. Brusaterra *et al.*, "Vertical GaN Trench MOSFETs With HfO2 / Al2O3 Layered Gate Dielectric," in *IEEE Transactions on Semiconductor Manufacturing*, https://ieeexplore.ieee.org/document/11121373

[3] M. Manasreh *et al.*, "A. Temperature Dependence of the Absorption Band Gap Edge of GaN," *MRS Online Proceedings Library* 395, 553–557 (1995). https://doi.org/10.1557/PROC-395-553

[4] J. Piprek *et al*., "Analysis of wavelength-dependent performance variations of GaN-based ultraviolet lasers." *Optoelectronic Devices: Physics, Fabrication, and Application IV*. Vol. 6766. SPIE, 2007

[5] M. Born *et al.*, "Principles of optics: electromagnetic theory of propagation, interference and diffraction of light." Oxford, Pergamon Press, 1964.

DOI 10.5162/iCCC2026/P45

Calculated electromigration wind force in molybdenum disilicide

Julia Baldauf
CiS Forschungsinstitut für Mikrosensorik GmbH, Erfurt, Deutschland
jbaldauf@cismst.de

Kurzfassung

Mit Hilfe von Machine Learning (random forest) konnte die Größe der Windkraft-Komponente der effektiven Ionenladung der Elektromigration in Molybdändisilizid für verschiedene Atomsorten bestimmt werden.
Es ist zu erkennen, dass alle Atomsorten einen vergleichsweise kleinen Wert von Z^*_{wind} ≈-1 für diese Komponente aufweisen. Da in der Regel die direkte Komponente einen wesentlich kleineren Beitrag zur effektiven Ionenladung ergibt als die Windkraft-Komponente, kann die hier bestimmte Windkraft-Komponente als Näherung der effektiven Ionenladung.

Abstract

Using a machine learning approach with a random forest model the value of the wind force component of the effective ion charge of electromigrating ions in molybdenum disilicide was estimated for several ion species. The value of Z^*_{wind} ≈-1 is small compared to effective ion charges in other host materials. Because of the direct component having a small contribution to the effective ion charge compared to the wind force, the wind force component of the effective ion charge can be used as approximation for the effective ion charge itself.

1 Introduction

Because of chip sizes decreasing and current densities increasing electromigration becomes more important.
Electromigration is the movement of atoms in an electrically conductive material caused by current through the material.
Molybdenum disilicide ($MoSi_2$) and electromigration therein are of great importance because of the widespread use of the material in macroscopic and microscopic heating devices.
$MoSi_2$ is often used in microheaters because of its high melting point and its CMOS compatibility. The high melting point of $MoSi_2$ is the reason it has been considered to be immune to electromigration for a long time. This is the reason electromigration in the material has not been investigated thoroughly, despite several researchers proving the existence of electromigration phenomena in $MoSi_2$ [1].

1.1 Electromigration

The electromigration force $\vec{F}$ on an ion caused by the electric field $\vec{E}$ can be described as
$\vec{F} = Z^* e\vec{E} = (Z^*_{dir} + Z^*_{wind})e\vec{E}$.
With Z^* being the effective ion charge and e being the elemental charge.
$Z^* = Z^*_{dir} + Z^*_{wind} = Z^*_{dir} + K/\rho$,
with K depending on the material but being independent of the temperature [2].
For most materials Z^*_{dir} is small compared to Z^*_{wind} [2]. This allows the use of Z^*_{wind} as an approximation for Z^*in these materials
$Z^* \approx Z^*_{wind} = K/\rho$.
Because of the experimental challenges of the determination of Z^* simulations using density functional theory or machine learning might give adequate results much faster.

2 Method

In this work a machine learning approach was performed using the Materials Simulation Toolkit for Machine Learning (MAST-ML) [3] to determine K for a plethora of atoms in $MoSi_2$. The path averaged K values of [2] have been used as dataset because of K being available for a multitude of atom combinations.
Care was taken not to have multiples of a data point and to exclude data points in which the crystal structure mentioned in [2] would not have matched the crystal structure of the feature generation of the elemental feature generator. This resulted in 139 data points being used. The elemental properties were constructed using the Materials Agnostic Platform for Informatics and Exploration (MAGPIE) approach.
The elemental feature generator has been used to generate features of the combination of impurity and host material to give composition averages, arithmetic averages, the minimum, the maximum, and the differences between the two of them.
In the first runs of the machine learning algorithm the selection of the input parameters of the sequential forward selection algorithm (SFS) showed the melting temperature being an important parameter. The elemental feature generator cannot give an adequate melting temperature of $MoSi_2$, therefore the melting temperature of the host material, the electromigrating ion and the arithmetic average thereof were added as features.
The data was cleaned using imputation of the median for missing values.
Normalization was done to standardize the data prior to fitting. After normalization the mean of each feature is 0 and the standard deviation of the feature is 1.

A random forest regressor was chosen using 125 estimators employing an ensemble model feature selector including SFS.
The feature learning curve (Figure 1) showes a decrease of the mean absolute error until the number of 7 features is met.

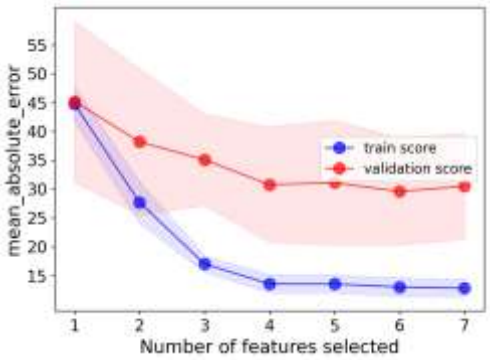


Figure 1 Feature learning curve.

Therefore, the number of 7 features was selected in the ensemble model feature selector. The features used are ranked according to their importance in Table 1.

Feature	Importance
IsMetalloid_difference	0.12795
Host-Melting-Temperature	0.09682
HeatFusion_max_value	0.05481
CovalentRadius_min_value	0.04667
ThermalConductivity_max_value	0.02988
NUnfilled_min_value	0.02681
Polarizability_min_value	0.02373

Table 1 Selected features and their importance.

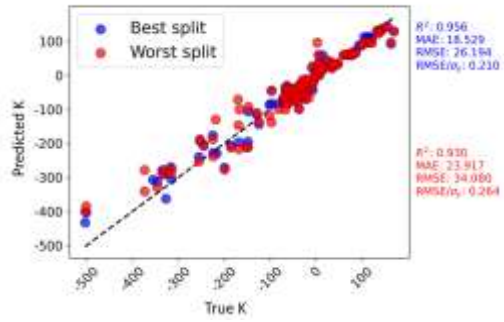


Figure 2 5-fold CV test.

The ρ values used in [2] to calculate Z^*_{wind} are the specific resistivities of polycrystalline materials at the temperature of 473 K.
For $MoSi_2$ a value of ρ=154.13 μΩ cm has been determined due to measurements of the resistance at 473 K and was used in the calculation of Z^*_{wind}.

Electromigrating material	Z^*_{wind}
Sc, Ti, V, Cr, Mn, Fe, Ga, Ge, Sr, Y, Zr, Nb, Mo, Tc, Ru, Cd, In, Sn, Ba, Hf, Ta, Re, Os, Hg, Tl, Pb, Ra, Rf, Db, Sg, Bh, Hs, La, Ce, Pr, Nd, Pm, Sm, Eu, Gd, Tb, Dy, Yb, Lu, Ac, Th, Pa, U, Np, Pu, Am, Cm, Bk, Cf, No, Lr	-1.0029
Li, Na, K, Rb, Cs, At, Fr, Tm, Md	-1.0141
He, B, Ne, Ar, Zn, Kr, Pd, Xe, Rn	-0.9885
Ni, Te, Pt, Po, Ds, Er, Fm	-0.9630
Co, Sb, Ir, Mt, Ho, Es	-1.0265
H, F, Cl, Br, I	-0.9985
P, As, Rh, Bi	-1.018
O, S, Se	-0.9553
Be, Ca	-1.0234
Ag	-1.1594
Au	-1.1398
Cu	-1.1338
Al	-1.066
Si	-1.0368
W	-1.0009
Mg	-0.9866

Table 2 Z^*_{wind} determined by machine learning of K.

3 Conclusion

The host melting temperature being one of the most important features influencing the effective ion charge is well known. The electrical conductivity is known to influence the effective ion charge. Instead of the electrical conductivity the SFS identified the thermal conductivity to be of importance. This is because of the strong correlation between the two of them.
The values of Z^*_{wind} being close to -1 is plausible.
Typical values of Z^* are either in the same range or one or two magnitudes larger.
The value of Z^*_{wind} being comparatively small in $MoSi_2$ might explain $MoSi_2$ to have been considered immune to electromigration.

4 Literature

[1] Zehe, A.; Ramirez, A.: Electromigration of aluminium through quasi bamboo-like grain blocked silicide interconnects, Crystal research and technology 35, 5 (2000), S. 557-562

[2] Dekker, J.P.; Lodder, A.: Calculated electromigration wind force in face-centered-cubic and body-centered-cubic metals, J. Appl. Phys 84 (1998), S. 1958-1962

[3] Jacobs, R., Mayeshiba, T., Afflerbach, B., Miles, L., Williams, M., Turner, M., Finkel, R., Morgan, D.: The Materials Simulation Toolkit for Machine Learning (MAST-ML): An automated open source toolkit to accelerate data- driven materials research, Comp. Mat. Sci. 175 (2020), 109544

DOI 10.5162/iCCC2026/P46

Quanteninterferenzmodule für die optische Kohärenztomographie zur zerstörungsfreien Prüfung keramischer Bauteile unter Zuhilfenahme „undetektierter“ Photonen im mittleren Infrarotbereich

Gunnar Blume[1], David Feise[1], Alexander Sahm[1], Nils Werner[1], Johannes Zender[1], Sven Ramelow[1,2], Katrin Paschke[1]
1 Ferdinand-Braun-Institut (FBH), Gustav-Kirchhoff-Straße 4, 12489 Berlin, Germany;
2 Institute for Physics, Humboldt-Universität zu Berlin, Newtonstraße 15, 12489 Berlin, Germany.

Kurzfassung

Der mittlere Infrarotbereich (MIR) bietet im Gegensatz zum sichtbaren oder nahinfraroten (NIR) Spektralbereich aufgrund geringerer Streuung die Möglichkeit zur zerstörungsfreien Prüfung von keramischen Bauteilen. Allerdings sind die Lichtquellen und Detektoren in diesem Bereich um Größenordnungen unempfindlicher. Daher wird die Quantenkorrelation zwischen Photonen im MIR und NIR benutzt, die mittels parametrischer Fluoreszenz in einem Interferometer-Aufbau erzeugt wurden. Wird dann mit den Photonen im MIR die Probe abgetastet, kann die Information aus dem Interferenzspektrum der NIR-Photonen gewonnen werden. Die MIR-Photonen bleiben dabei „undetektiert“. Bisherige Laboraufbauten wurden jetzt so weit miniaturisiert, dass sie in industrieller Umgebung eingesetzt werden können.

Abstract

The non-destructive testing of ceramics is preferably done in the mid-infrared (MIR) spectral region rather than the visible or near infrared (NIR) due reduced scattering. However, MIR light sources and detectors feature a significantly lower sensitivity than their NIR counter parts. Quantum correlation between MIR and NIR photons can be used to circumnavigate the challenge. Entangled photons are generated in a spontaneous parametric down conversion process, which is put in an interferometer setup. Thus, the MIR photon can be used to scan a sample, while the NIR photons which also carry the same interference information due to the correlation are detected. The MIR photons remain “undetected” in the whole process. The setup was now miniaturized enabling future industrial application.

1 Einleitung

Für viele Fertigungsprozesse ist ein zerstörungsfreier Messprozess wünschenswert. Optische Messungen werden in der automatisierten Produktion in großer Zahl verwendet, da hier zusätzlich kein physischer Kontakt zum Messobjekt notwendig ist (siehe Bild 1). Die Wahl der verwendeten Wellenlänge spielt dabei eine entscheidende Rolle. Bei Keramiken hat z.B. eine Wellenlänge im mittleren Infrarot (MIR) den Vorteil der geringeren Streuung und damit größeren Eindringtiefe. Allerdings sind die Leistungen der Quellen und die Effizienz der Detektion um Größenordnungen geringer als im sichtbaren oder nahinfraroten (NIR) Bereich.
Eine elegante Methode um im MIR messen und im NIR detektieren zu können, ist die Nutzung der Quantenkorellation zwischen Photonen im MIR und NIR. Ein solches korreliertes (und verschränktes) Photonenpaar lässt sich mit parametrischer Fluoreszenz (spontaneous parametric down conversion, SPDC) in einem passenden nichtlinearen Kristall erzeugen [1,2,3].

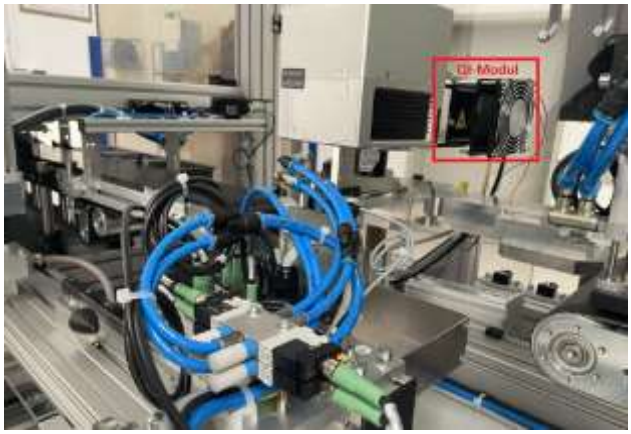


Bild 1 Foto eines Quanteninterferenzmoduls im Messaufbau in einer Sortieranlage (mit freundlicher Genehmigung von NELA).

2 Konzept

Um ein Interferenz-Messsignal zu erhalten, ist es notwendig, dass sowohl MIR- als auch die NIR-Photonen sowie das Pumplicht zweimal einen SPDC-Prozess durchlaufen

DOI 10.5162/iCCC2026/P46

(siehe Bild 2). Hierzu wird ein nichtlinearer Kristall (periodisch gepoltes Kaliumtitanylphosphat, ppKTP) mit einer Wellenlänge von 660 nm optisch gepumpt. Der Pumplaser ist in unserem Fall ein Halbleiterlaser, der extra für diesen Einsatz entwickelt wurde, um ausreichend kohärentes Pumplicht bei kleinsten Abmaßen bereitzustellen [4].

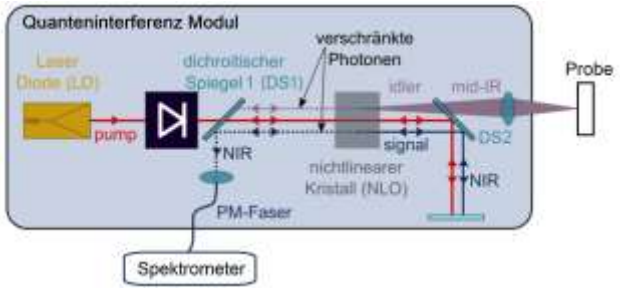


Bild 2 Prinzipskizze des Messens mit „undetektierten" Photonen

Das Pumplicht passiert einen optischen Isolator und wird geeignet auf den ppKTP-Kristall fokussiert, um verschränkte NIR- (um ca. 800 nm) und MIR-Photonen (um ca. 3.8 µm) zu erzeugen [2]. Danach wird das NIR-Licht und die Pumpstrahlung vom MIR-Anteil getrennt (Langpassfilter, DS 2) und verbleibt auf einer Referenzstrecke im Quanteninterferenz-Modul. Die MIR-Photonen verlassen das Modul, werden auf eine Probe gelenkt, von der sie zurückgestreut oder reflektiert werden und auf dem Rückweg ein zweites Mal den ppKTP Kristall passieren. Das geschieht zeitgleich zusammen mit dem NIR und Pumplicht aus dem Referenzarm. Dadurch können alle Photonen miteinander interferieren (Quantenvariante eines Michelson-Interferometers). Dabei entsteht ein Interferenzsignal nicht nur in den MIR-Photonen sondern auch in den korrelierten NIR-Photonen. Letztere werden dann über einem dichroitischen Strahlteiler (Kurzpassfilter, DS1) in eine optische Faser gekoppelt und zu einem Spektrometer mit einer Silizium-Detektorzeile geleitet.

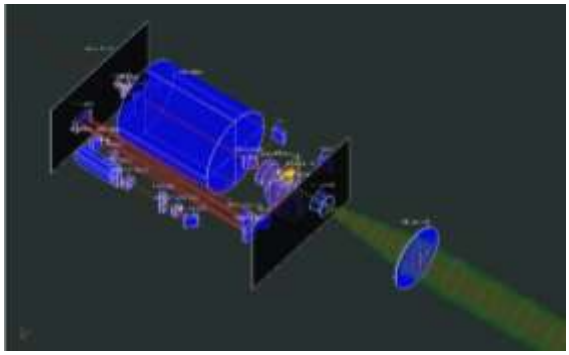


Bild 3 Optische Simulation des Strahlengangs im Quanteninterferenzmodul.

Für die Miniaturisierung des Aufbaus wurde als erstes der optische Strahlengang mit Hilfe der Software BeamXpert DESIGNER simuliert (siehe Bild 3). Besondere Aufmerksamkeit wurde auf die Dispersionseigenschaften der verwendeten Optiken gelegt, um möglichst viele NIR-Photonen detektieren zu können. Daraus wurde ein mechanisches Modell entwickelt, welches neben den Optiken auch elektrische Leiterbahnen, eine interne Peltierkühlung integriert sowie optischen und elektrische Anschlüsse ermöglicht. Das Gehäuse hat dabei eine Grundfläche von 75 x 90 mm² (siehe Bild 4).

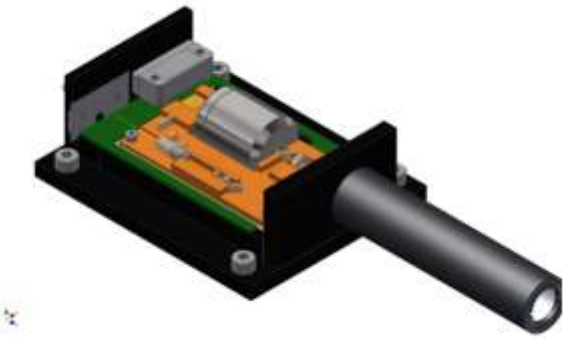


Bild 4 Renderbild der mechanischen Zeichnung des Quanteninterferenzmoduls.

Bau und Messergebnisse

Daraufhin wurden alle Einzelteile angefertigt (Optikträger Inlay, Laser, Gehäuse) bzw. kommerziell erworben (optischer Isolator, Mikrooptiken, Photodioden, Temperatursensoren, Faser und Elektroanschlüsse). Mittels Mikromontage wurden sämtliche Komponenten im Modul mit einer Genauigkeit von teilweise < 0.2 µm platziert und montiert (siehe Bild 5).

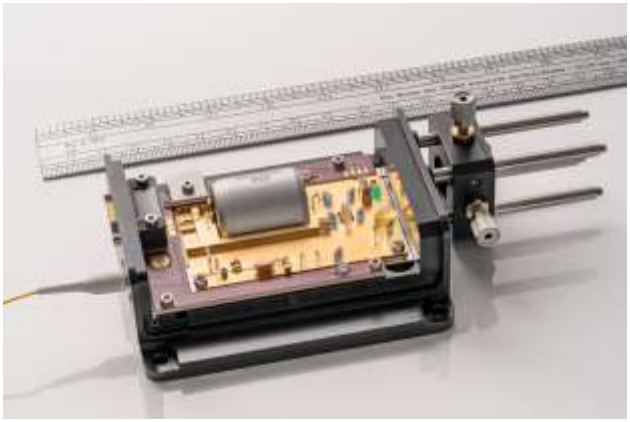


Bild 5 Foto eines Quanteninterferenzmoduls mit Mikrooptiken und Maßstab zum Größenvergleich.

Zusätzlich wurde eine externe Elektronik entwickelt, die eine Leistungsstabilisierung des Pumplasers ($\Delta P < 1\%$ über 1 h) und eine Temperaturstabilisierung des gesamten Optikträgers (± 0.2 K) ermöglicht [5].
Mit dem Quanteninterferenzmodul und einem Spektrometer (HR-6N750-50WG6) mit konventionellem Silizium-Zeilendetektor konnte dann das NIR-Licht mit einer Gesamtphotonenrate von ca. $10^8 s^{-1}$ detektiert werden. Wird ein Silberspiegel als Probe für das mit einer externen ZnSe-Linse kollimierte MIR-Licht benutzt, zeigt sich im NIR-Licht Interferenz mit einer Sichtbarkeit von ca. 39%.

DOI 10.5162/iCCC2026/P46

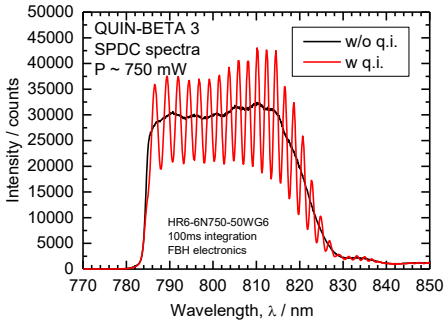


Bild 6 Spektren des detektierten nahinfraroten Anteils der parametrischen Fluoreszenz ohne (schwarz) und mit (rot) Quanteninterferenz

Wird das SPDC-Licht ohne Quanteninterferenz vom Signal mit aufmodulierter Quanteninterferenz abgezogen, ergibt sich ein reines Interferenzsignal dessen Frequenz sich aus dem Unterschied der Weglängen zwischen dem MIR-Licht und der NIR-Referenzstrecke ergibt (siehe Bild. 7).

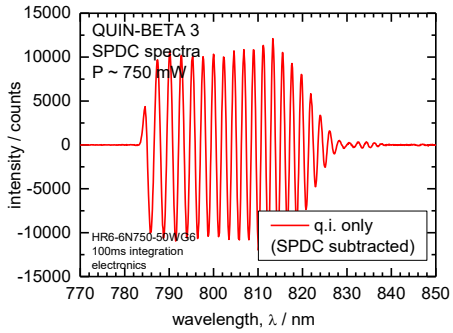


Bild 7 Quanteninterferenz abzüglich des Untergrundes der parametrischen Fluoreszenz

Um einen Anwendungsfall zu demonstrieren, wurden drei verschiedene CaF_2-Spiegel, die eine hochreflektierende Schicht im sichtbaren und NIR-Bereich besitzen, als Proben mit MIR-Photonen untersucht. Die CaF_2-Spiegel haben Dicken von 1 mm, 2 mm und 3 mm (siehe Bild 8).

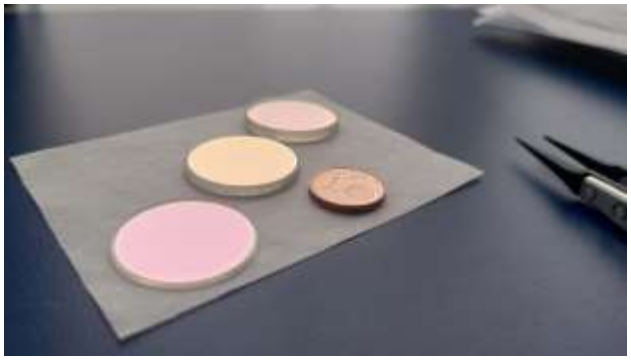


Bild 8 Foto von drei CaF_2 Proben mit d = 1 mm, 2 mm und 3 mm. Die Proben sind hochreflektierend im sichtbaren und nachinfraroten Bereich.

Werden die Proben in den MIR Strahl gehalten, verlängert sich der optische Weg um die doppelte Dicke mal dem Brechungsindex (n = 1.4). Das verändert die Frequenz des Interferenzsignals (siehe Bild 9).

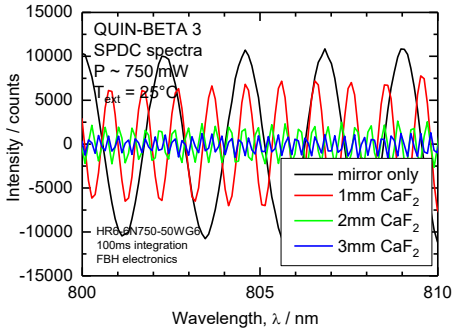


Bild 9 Ausschnitt aus dem Quanteninterferenzspektren für verschieden dicke CaF_2-Proben.

Dabei ändert sich allerdings nicht nur die Frequenz, sondern durch Absorption und Streuung auch die Amplitude. Um die Größen besser darzustellen wurde an den Spektren eine schnelle Fourier-Transformation (FFT) durchgeführt (siehe Bild 10).

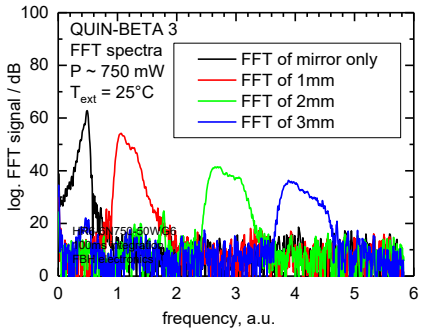


Bild 10 Fast Furier Transformierte (FFT) der QI-Spektren aus denen sich, bei bekanntem Brechungsindex, direkt die Dicke der Proben ableiten lässt.

Aus Bild 10 wird ersichtlich, dass alle drei Proben vermessen werden konnten und diese die Frequenz der Quanteninterferenz um den Faktor ihrer Dicke ändern. Im Fall der 3 mm dicken Probe ist allerdings eine Signalabschwächung von ca. 25 dB im Verhältnis zum rein am Silberspiegel reflektierten MIR Licht zu beobachten.

3 Zusammenfassung

Es wurden Quanteninterferenzmodule entwickelt, die es ermöglichen im industriellen Umfeld optische Dickenmessungen im mittleren Infrarot (MIR) durchzuführen. Dabei bleiben die eigentlichen MIR-Photonen „undetektiert“. Stattdessen werden in einem Interferometeraufbau verschränkte Photonen im nahen Infrarot detektiert. Die NIR Photonen können unkompliziert und kostengünstig mit einem Si-basierten Spektrometer detektiert und daraus die Quanteninterferenzinformation gewonnen werden. Somit

erhält man mit diesen Modulen einen anwenderfreundlichen Zugang zu Spektren und Tiefeninformation im mittleren Infrarot, was zukünftig die Messtechnik und Produktionsüberwachung beispielsweise von Keramikenbauteilen vereinfacht.

4 Danksagung

Die Autoren bedanken sich beim BMFTR für die Förderung des Projektes QUIN (FKZ 13N15402, 13N15403) und der Forschungsfabrik Mikroelektronik Deutschland (FMD). Weiterer Dank gilt der Firma NELA Brüder Neumeister GmbH für die gute Zusammenarbeit und Tests unter Industriebedingen.

5 Literatur

[1] Lemos, G. B., Borish, V., Cole, G. D., Ramelow, S., Lapkiewicz, R., Zeilinger, A. "Quantum imaging with undetected photons," Nature 512, 409–412 (2014). https://doi.org/10.1038/nature13586

[2] Vanselow, A., Kaufmann, P., Zorin, I., Heise, B., Chrzanowski, H.M., and Ramelow, S., "Frequency-domain optical coherence tomography with undetected mid-infrared photons," Optica 7 (12), pp. 1729-1736 (2020). https://doi.org/10.1364/OPTICA.400128

[3] Pearce, E., Gemmell, N.R., Flórez, J., Ding, J., Oulton, R.F., Clark, A.S., Phillips, C.C., "Practical quantum imaging with undetected photons," Optics Continuum 2 (11), 2770-0208 (2023). https://doi.org/10.1364/OPTCON.507154

[4] Blume, G., Matalla, O., Wenzel, H., Maaßdorf, A., Feise, D., Fricke, J., Ressel, R., Kreutzmann, S., Ginolas, A., Sahm, A., Knigge, A., Paschke, K., "Single mode 660 nm DBR tapered laser with 1 W optical output power," Proc. of SPIE, 12402, 1240205 (2023). https://doi.org/10.1117/12.2647842

[5] Blume, G., Feise, D., Sahm, A., Hildenstein, P., Werner, N., Zender, J., Wendt, F., Gewers, F., Pearce, E., Ramelow, S., Paschke, K., "Long wavelength, small quantum interferometer module: shrinking MIR OCT systems via quantum entanglement," Proc. of SPIE 13392, 133920V (2025). https://doi.org/10.1117/12.3044313

DOI 10.5162/iCCC2026/P47

THE IMPORTANCE OF PREDICTING BATTERY DEGRADATION IN LIGHT ELECTRIC AIRCRAFT

V.P. Joshi, R. Cepeda-Gomez, D. Knieschke
German Aerospace Center (DLR), Institute of Electrified Aero Engines, Department of Aeronautical Requirements and Control of Propulsion System, Cottbus, Germany

vimarsh.joshi@dlr.de, rudy.cepedagomez@dlr.de, daniel.knieschke@dlr.de

Abstract

This paper discusses the importance and need for predicting battery degradation in an all-electric battery powered aircraft. This type of aircraft is expected to represent an emission-free alternative to the conventional light aircraft used for pilot training and leisure flying. The drawbacks of limited life-span of batteries, important parameters like total charge capacity and internal resistance which indicate battery health, and the need to define end-of-life (EOL) parameter values are addressed. The importance of analysing the particularities of the mission flown by light planes and the different approaches to model battery cells for battery management system (BMS) algorithms are presented. Further avenues to continue the work are also presented.

1. INTRODUCTION

The aviation industry is responsible for nearly 2% of the global CO_2 and 3% of the global NO_x emissions. This will increase in the foreseeable future together with the demand for air travel [1]. Electric propulsion technology has the potential to curb aircraft emissions and achieve the climate goals set by the commercial aviation industry and different political agendas of the governments and agencies [2]. All-electric propulsion architectures powered solely by batteries produce zero emissions during flight operations. In [3], Pattanyak et al. conducted a comprehensive study reviewing different battery technologies for aviation and concluded that for the foreseeable future lithium-ion (Li-ion) cell chemistries will dominate the battery technology required for electric aviation considering their high gravimetric and volumetric energy and power densities, low self-discharge rate, lower degradation rate, etc. However, the gravimetric and volumetric energy and power density of state-of-the-art Li-ion chemistry is still relatively low compared to conventional fuels or hydrogen. This makes them unpractical for long distance transport aircraft [4]. Electric power trains with Li-ion battery pack are nevertheless suitable for light aircraft used for leisure flying and pilot training such as one shown in Fig. 1.

Such light battery-powered aircraft can reduce the environmental footprint and operating cost of flight school, especially when the electricity used to charge the batteries is generated from renewable energy sources [6].

One of the main challenges in designing battery-powered aircraft is to predict how long the battery pack will last. The knowledge of useful battery life is particularly important from the perspective of certification. With time, the energy and power capacity of a battery pack degrades due to calendar and cycle aging [7]. This prompts the need to accurately monitor battery health, predict degradation, and optimize operation over the whole life cycle to increase the useful life and reliability of the battery pack without compromising the flight safety. The task is not trivial as the aging and degradation of Li-ion batteries is a complex non-linear process and predicting degradation dynamics is challenging [8]. This can be accomplished by intelligent battery management system (BMS) algorithms similar to those used in the automotive sector.

This work explores some important considerations in this regard and proposes an approach on how to address this issue.

2. BATTERY AS A LIFE-LIMITED COMPONENT

Several components and systems in aviation have a limited life span including batteries. As battery cells undergo ageing and degradation, their total charge capacity decreases and internal resistance increases. This happens because of the unwanted side reactions that consume lithium ions and structural deterioration of the electrodes. Thus, they last a limited number of cycles or operational hours before having to be replaced. Batteries are expensive: in mobility applications like electric vehicles (EVs), the battery system is one of the major contributors to the overall cost of an EV, reaching about 40-60% of the total cost price of the vehicle [9]. Study of battery degradation is also very important from the perspective of aircraft operation eco-

FIG 1. Diamond eDA40 aircraft [5]

DOI 10.5162/iCCC2026/P47

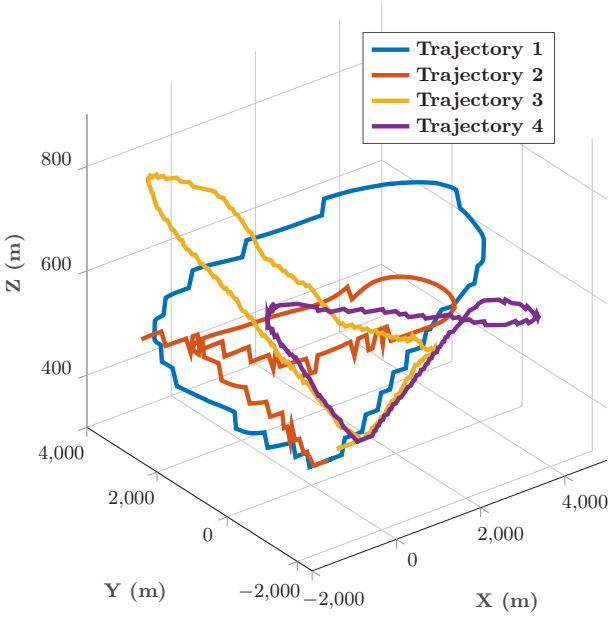


FIG 2. Flight trajectories of a light aircraft [11]

nomics. Thus, there is a need to determine its initial use end-of-life point in a better way.

The common practice followed in the automotive industry is to replace the battery pack once its energy storage capacity degrades to 70-80% of the value at the beginning-of-life (BOL). This degradation to 80% of the BOL energy storage capacity is called the end-of-life (EOL) point for a battery pack. Saxena et al. [10] showed that batteries in electrical vehicles continue to meet the performance requirements of the daily user even when energy storage capacity has degraded to a value well below 80% and the power delivering capacity has been degraded down to 30%. They conclude that accepted practice of retiring batteries when energy storage capacity has degraded to 80% is inaccurate, and should be done when the battery pack no longer meets the user requirement. This retired battery might then be used for a second-life stationary application.

This prompts the need to investigate and define the EOL parameters for the battery pack in a light electric aircraft as the dynamics of load profile and operating conditions are very different compared to automotive and it is a mandatory requirement for certification.

3. EFFECTS OF THE MISSION PROFILE

To investigate its degradation, it is important to identify the operating conditions which the battery pack will be subjected to during its life cycle. As mentioned in 1, such light electric aircraft can be used for leisure flying and pilot training, and their mission profile can vary a lot. This is unlike commercial aircraft, which usually fly similar profiles during their whole operational life. During flight training missions, aircraft usually remain in the vicinity of the airfield doing traffic pattern training. This leads to very frequent cycles between idle and full power. While flying, it can encounter stochastic conditions like turbulence, emergency maneuvers, etc. Several flights can be flown during a day which might require repeated charging of batteries. All these factors can affect battery performance and health.

Obtaining load profiles for a light electric aircraft is challenging because operational data of such aircraft is not publicly available. Some information about mission trajectories of light conventional aircraft can be extracted from data used for aircraft surveillance, like Automatic Dependent Surveillance Broadcast (ADS-B) or Mode S. In such surveillance technique, aircraft automatically transmit their identity, position, speed, etc., derived from onboard systems to other aircraft in the vicinity and ground air traffic control stations to avoid mid-air collision, and for air traffic monitoring and management [12].

Many datasets are publicly available and can be used for generating and analyzing flight trajectory patterns. One such trajectory dataset for light general aviation (GA) was collected and made available by Patrikar, et al. using an on-site ADS-B receiver at Pittsburgh-Butler Regional Airport (ICAO location indicator KBTP), near the city of Pittsburgh in Pennsylvania, USA [11]. Few mission trajectories flown by light conventional aircraft are shown in Fig. 2. It can be seen that even though the aircraft take off and land on the same airfield, their mission trajectories can be very different. For example, trajectories 1 and 2 have steps in the flying pattern indicating that multiple climb, cruise and descent maneuvers with steep pitch angles were performed during the flight. Such maneuvers have an impact on the power drawn from the battery pack and this needs to be taken into consideration when defining the constraints to optimize battery pack design, performance, and life.

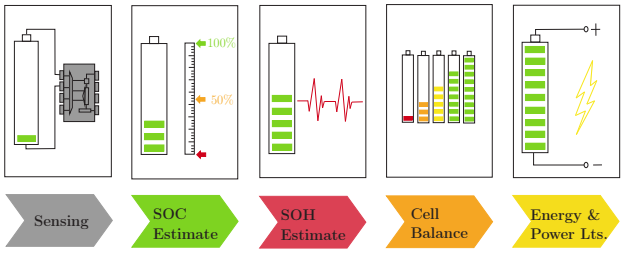


FIG 3. BMS Workflow [13]

4. BATTERY CELL MODELS FOR BMS ALGORITHMS

As described in section 1, the task of predicting battery degradation is not trivial and it is accomplished by intelligent BMS algorithms. Values of voltage, temperature, and current from each cell are measured in real-time using appropriate sensors and communicated to the BMS. These values are used to estimate the state-of-charge (SOC), state-of-health (SOH), perform cell balancing, and compute energy and power limits. The workflow of BMS is shown in Fig. 3. Battery management algorithms perform these estimations using observers which are mathematical models of the battery cells. SOC and SOH estimations are performed for each cell in the pack. Based on the number of cells connected in series and parallel, cell balancing is performed, and then energy and power limits of the pack are computed. The cell models used for such estimations can be physics-based, data-driven, or a combination of both.

SOH is indicated by analyzing two parameters: total charge capacity and internal resistance of the cells. They provide an estimate of the total energy storage and power delivering capacities. As the battery pack undergoes aging and degradation, the total charge capacity and the internal resistance of the cells change. Obtaining a good estimate of these values is crucial [14]. BMS SOH algorithms accomplish this with the help of observers. This makes an accurate cell model necessary.

DOI 10.5162/iCCC2026/P47

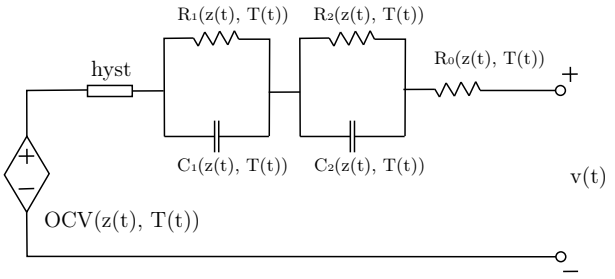


FIG 4. ECM

Physics-based models are developed using the partial differential equations that govern various electrochemical reactions occurring in the electrodes and electrolytes of the cell. Such models have higher fidelity, as they aim to capture all key behaviours of the cell to provide high accuracy, but they are very expensive computationally [15]. This renders them unsuitable for real-time applications.

The second category of models are data-driven models in which experimental data is used to parameterize a predefined model structure and validate behaviour. These models depend on the design of the experiments and quality of the data available. In this category there are two subtypes. The first one uses experimental data to fit empirically derived equations to infer input-output relationships. Such models are simple and easy to implement, but they are of low fidelity and have high inaccuracies. It is also difficult to interpret their parameters physically. These are also known as black-box models. This category also includes Artificial Intelligence based models which have become very popular in the recent years because of their better performance compared to conventional black-box models but they are not considered in this study because of the limitations in certifying them according to stringent aviation standards. The second category consists of lumped-parameter equivalent circuit models (ECM). These models use different electrical circuit elements to mimic the behaviour of an actual battery cell. The simplest model consists of an ideal voltage source connected in series to a resistance. Such models are static in nature and used for preliminary sizing and design studies. Higher fidelity ECMs take into account the dynamic effects by connecting one or more parallel RC networks or some other circuit elements in series or parallel to the series resistance and voltage source. These additional elements characterize the dynamic behaviour of various electrochemical processes happening inside a cell. Such ECMs have found traction on BMS applications since they offer a good trade-off between accuracy and computational effort, while maintaining a strong physical correlation between the model parameters and actual electrochemical processes within the battery cell [16].

Many comprehensive studies have been conducted to compare different ECMs. Hu, et al. did a comparative study of 12 ECMs for Li-ion battery cells in [15]. In [16], Nejad, et al. compared 7 different types of ECMs for real-time estimation of Li-ion battery states. In [17] and [18] different ECMs were investigated for their application to EV. In [19], Tran et al. developed a comprehensive ECM to incorporate effects of SOH, SOC, and temperature on model parameters.

Based on the literature review, a dual polarization ECM with one hysteresis element was selected for further study. The model structure is shown in Fig. 4. It is important to include the hysteresis element in the model to avoid

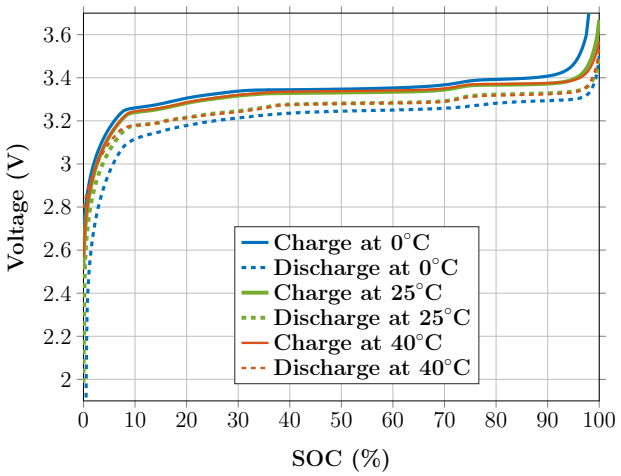


FIG 5. OCV vs SOC curves [20]

Temperature (°C)	OCV RMSE (V)
-10	0.2132
0	0.1980
20	0.1319
25	0.0638
30	0.1282
40	0.0614
50	0.2244

TAB 1. RMSE between charge and discharge OCV curves at different temperatures [20]

SOC estimation errors. Hysteresis is a phenomenon which causes the difference in the charge and discharge OCV curves for same SOC. This phenomenon is shown in Fig. 5 using experimental data of A123 Li-ion cell collected by by Xing, et al. which is publicly available. Table 1 summarizes the root mean square error between charge and discharge OCV-SOC data for different temperatures.

5. NEXT STEPS

The preliminary discussion presented in this paper will be elaborated further in the future. The main avenues of development are a more informed determination of the expected operational profiles of a light aircraft. This will be done by detailed analysis of open-source ADS-B data of light aircraft and consultation with the operators of light aircraft (flight schools, flight clubs) and manufacturers of some of the light electric airplanes already certified or about to be. Once the operational profiles of light aircraft are identified, they will be used to generate synthetic cell cycling data to parameterize the ECM proposed in this paper. The parameterized ECM will then be used to develop BMS algorithms for predicting battery pack degradation to optimize its operation with objective to increase battery life.

6. CONCLUDING REMARKS

This paper discussed the importance of predicting battery degradation during the operational life of a batterypowered light aircraft. It was discussed how the individual characteristics of the different operational profiles of this class of airplane represent a particular challenge in the safe

and reliable operation of the batteries. The importance of modeling battery cells to create BMS algorithms used for estimating battery health and predicting degradation was highlighted since this could maximize the useful life of the battery. This study also explored the different types of cell models which could be used as observers in BMS algorithms and selected a model relevant to the use application presented in this study.

References

[1] J. Mukhopadhaya and B. Graver. Performance analysis of regional electric aircraft. *International Council of Clean Transportation white paper*, 2022.

[2] V.P. Joshi, R. Cepeda-Gomez, and T. F. Geyer. Optimal sizing and energy management of a battery to support a fuel cell powered regional electric aircraft during take-off phase. In *AIAA AVIATION FORUM AND ASCEND*, 2024. DOI: 10.2514/6.2024-3702.

[3] Tavish Pattanayak and Dimitri Mavris. Battery technology for sustainable aviation: a review of current trends and future prospects. *Applied Energy*, 397:126356, 2025. ISSN: 0306-2619. DOI: https://doi.org/10.1016/j.apenergy.2025.126356.

[4] A. T. Isikveren, A. Seitz, P. C. Vratny, C. Pornet, K. O. Plötner, and M. Hornung. Conceptual studies of universally-electric systems architectures suitable for transport aircraft. *DeutscherLuft und RaumfahrtKongress, DLRK, Berlin*, 2012.

[5] Diamond Aircraft Industries, 2025. https://www.diamondaircraft.com/en/flight-school-solution/aircraft/eda40/overview/.

[6] Nicolò Avogadro and Renato Redondi. Demystifying electric aircraft's role in aviation decarbonization: Are first-generation electric aircraft cost-effective? *Transportation Research Part D: Transport and Environment*, 130:104191, 2024. ISSN: 1361-9209. DOI: https://doi.org/10.1016/j.trd.2024.104191.

[7] Yaqi Li, Wendi Guo, Daniel-Ioan Stroe, Hongbo Zhao, Peter Kjær Kristensen, Lars Rosgaard Jensen, Kjeld Pedersen, and Leonid Gurevich. Evolution of aging mechanisms and performance degradation of lithium-ion battery from moderate to severe capacity loss scenarios. *Chemical Engineering Journal*, 498:155588, 2024. ISSN: 1385-8947. DOI: https://doi.org/10.1016/j.cej.2024.155588.

[8] Ruihe Li, Niall Kirkaldy, Fabian Oehler, Monica Marinescu, Gregory Offer, and Simon O'Kane. The importance of degradation mode analysis in parameterising lifetime prediction models of lithium-ion battery degradation. *Nature Communications*, 16, 03 2025. DOI: 10.1038/s41467-025-57968-3.

[9] H. Proff. Study: The key role of battery costs in automotive, 2023. https://www.deloitte.com/de/de/Industries/automotive/research/study-key-role-of-battery-costs-in-automotive.html.

[10] Samveg Saxena, Caroline Le Floch, Jason MacDonald, and Scott Moura. Quantifying ev battery end-of-life through analysis of travel needs with vehicle powertrain models. *Journal of Power Sources*, 282:265–276, 2015. ISSN: 0378-7753. DOI: https://doi.org/10.1016/j.jpowsour.2015.01.072.

[11] Jay Patrikar, Brady Moon, Sourish Ghosh, Jean Oh, and Sebastian Scherer. TrajAir: A General Aviation Trajectory Dataset. 6 2021. DOI: 10.1184/R1/14866251.v1.

[12] Junzi Sun, Joost Ellerbroek, and Jacco Hoekstra. Flight extraction and phase identification for large automatic dependent surveillance–broadcast datasets. *Journal of Aerospace Information Systems*, 14:1–6, 08 2017. DOI: 10.2514/1.I010520.

[13] Munish Manas, Ravish Yadav, and Rajesh Kumar Dubey. Designing a battery management system for electric vehicles: A congregated approach. *Journal of Energy Storage*, 74:109439, 2023. ISSN: 2352-152X. DOI: https://doi.org/10.1016/j.est.2023.109439.

[14] Gregory L. Plett. Extended kalman filtering for battery management systems of lipb-based hev battery packs: Part 2. modeling and identification. *Journal of Power Sources*, 134(2):262–276, 2004. ISSN: 0378-7753. DOI: https://doi.org/10.1016/j.jpowsour.2004.02.032.

[15] Xiaosong Hu, Shengbo Li, and Huei Peng. A comparative study of equivalent circuit models for li-ion batteries. *Journal of Power Sources*, 198:359–367, 2012. ISSN: 0378-7753. DOI: https://doi.org/10.1016/j.jpowsour.2011.10.013.

[16] S. Nejad, D.T. Gladwin, and D.A. Stone. A systematic review of lumped-parameter equivalent circuit models for real-time estimation of lithium-ion battery states. *Journal of Power Sources*, 316:183–196, 2016. ISSN: 0378-7753. DOI: https://doi.org/10.1016/j.jpowsour.2016.03.042.

[17] Xiaofeng Ding, Donghuai Zhang, Jiawei Cheng, Binbin Wang, and Patrick Chi Kwong Luk. An improved thevenin model of lithium-ion battery with high accuracy for electric vehicles. *Applied Energy*, 254:113615, 2019. ISSN: 0306-2619. DOI: https://doi.org/10.1016/j.apenergy.2019.113615.

[18] Merve Tekin and M. İhsan Karamangil. Comparative analysis of equivalent circuit battery models for electric vehicle battery management systems. *Journal of Energy Storage*, 86:111327, 2024. ISSN: 2352-152X. DOI: https://doi.org/10.1016/j.est.2024.111327.

[19] Manh-Kien Tran, Manoj Mathew, Stefan Janhunen, Satyam Panchal, Kaamran Raahemifar, Roydon Fraser, and Michael Fowler. A comprehensive equivalent circuit model for lithium-ion batteries, incorporating the effects of state of health, state of charge, and temperature on model parameters. *Journal of Energy Storage*, 43:103252, 2021. ISSN: 2352-152X. DOI: https://doi.org/10.1016/j.est.2021.103252.

[20] Yinjiao Xing, Wei He, Michael Pecht, and Kwok Leung Tsui. State of charge estimation of lithium-ion batteries using the open-circuit voltage at various ambient temperatures. *Applied Energy*, 113:106–115, 2014. ISSN: 0306-2619. DOI: https://doi.org/10.1016/j.apenergy.2013.07.008.

DOI 10.5162/iCCC2026/P48

Lastanalyse zur Energieprognose einer universitären Wechselstrom-Ladeinfrastruktur für Elektrofahrzeuge

Tom Scheppan, *FG Dezentrale Energiesysteme und elektrische Netze – BTU Cottbus-Senftenberg,* Cottbus – Deutschland, tom.scheppan@b-tu.de
Tine Fischer, *FG Hochspannungstechnik und elektrische Anlagen – BTU Cottbus-Senftenberg,* Cottbus – Deutschland, tine.fischer@b-tu.de
Dirk Lehmann, *FG Hochspannungstechnik und elektrische Anlagen – BTU Cottbus-Senftenberg,* Cottbus – Deutschland, dirk.lehmann@b-tu.de
Mario Schenk, *FG Hochspannungstechnik und elektrische Anlagen – BTU Cottbus-Senftenberg,* Cottbus – Deutschland, mario.schenk@b-tu.de
Kaveh Malekian, *FG Dezentrale Energiesysteme und elektrische Netze – BTU Cottbus-Senftenberg,* Cottbus – Deutschland, malekian@b-tu.de

Kurzfassung

Die Studie untersucht 1.339 Ladevorgänge eines universitären AC-Ladeparks (28.06.2024 – 04.08.2025) zur Verbesserung der Lastprognosen und des Energiemanagements. Das Ladeverhalten folgt klar den Arbeitszeiten, mit Spitzen zwischen 6:30 und 11:00 Uhr. Pro Tag treten meist 1 – 6 Ladevorgänge auf, maximal 11. Das Wetter hat keinen messbaren Einfluss, während der Schulferien in Berlin/Brandenburg reduziert sich der Energiebedarf um ca. 22 %. Die homogene Nutzergruppe aus regelmäßig pendelnden Mitarbeitenden erzeugt stabile, übertragbare Lastprofile. Mit zunehmenden Batteriekapazitäten und künftigem bidirektionalem AC-Laden werden sich diese Profile jedoch verändern, wodurch flexible, intelligente Ladestrategien erforderlich werden.

Abstract

This study analyzes 1,339 charging sessions from a university AC charging park (28 June 2024 – 4 August 2025) to improve load forecasting and energy management. Charging behavior closely follows employees' working hours, with peaks between 6:30 and 11:00 a.m. Most days show 1 – 6 charging sessions, with a maximum of 11. Weather conditions have no measurable impact, while vacation periods reduce energy demand by about 22 %. The homogeneous user group of regular commuters results in stable and transferable load profiles. However, increasing battery capacities and the introduction of bidirectional AC charging will alter these patterns, highlighting the need for flexible and intelligent charging strategies.

1 Einführung

Die Elektrifizierung des Verkehrssektors führt zu steigenden Anforderungen an Verteilnetze und Ladeinfrastruktur. Insbesondere Elektrofahrzeuge bieten jedoch gleichzeitig im Kontext zukünftiger Vehicle-to-Grid-(V2G)-Technologien flexible Einsatzpotenziale für Energiemanagementstrategien. [1, 2] Der universitäre Ladepark der BTU bildet durch eine konstante, homogene Nutzerschaft eine geeignete Möglichkeit zur Analyse. Zwischen Juni 2024 und August 2025 wurden sämtliche AC-Ladevorgänge aufgezeichnet und hinsichtlich zeitlicher, leistungsbezogener und nutzungsbezogener Verteilungen, untersucht. Die Analyse soll Grundlage für Prognosemodelle und ein optimiertes, sektorgekoppeltes Energiemanagement des MicroGrids bilden.

2 Systemüberblick

Die Anlage umfasst 15 AC-Ladepunkte mit je 22 kW, deren Gesamtleistung über ein dynamisches Lastmanagement auf 90 kW begrenzt ist. Der Ladepark ist vollständig in das universitäre MicroGrid integriert, welches neben einer 99kVA PV-Anlage auch einen Batteriespeicher, ein Blockheizkraftwerk sowie weitere thermische Anlagen umfasst. Alle Anlagen werden von einem zentralen Leitsystem gesteuert.

3 Methodik

Die Datengrundlage bilden 1.339 Ladevorgänge, deren Startzeit, geladene Energie und Ladeleistung erfasst wurden. Die AC-Ladevorgänge werden durch eine konstante Leistung über die gesamte Ladedauer modelliert [3]. Alle Nutzer werden als Mitarbeitende mit regelmäßigen Campus-Routinen betrachtet. Wetter und variable Stromtarife wurden aufgrund eines Flatrate-Tarifs als nicht verhaltensbestimmend angenommen. Es wurden zeitliche Verteilungen, Leistungsprofile, Nutzerindividuen sowie externe Faktoren (Wetter, Ferien, Wochenenden) analysiert. Pearson- und Spearman-Korrelationen bewerten Zusammenhänge zwischen Energienachfrage und meteorologischen Größen.

DOI 10.5162/iCCC2026/P48

4 Ergebnisse

4.1 Analyse Ladevorgänge

Die Startzeiten der Ladevorgänge zeigen eine Häufung am Morgen. Zwischen 6:30 und 11:00 Uhr ist die Wahrscheinlichkeit von Ladestarts ansteigend, wobei das Maximum zwischen 8:30 und 8:45 Uhr liegt. Nach 11:00 Uhr sinkt die Wahrscheinlichkeit neuer Ladevorgänge signifikant. Pro Tag werden meist 1 – 6 Ladevorgänge registriert, 10 – 11 Ladevorgänge an einem Tag sind seltene Ausnahmen.

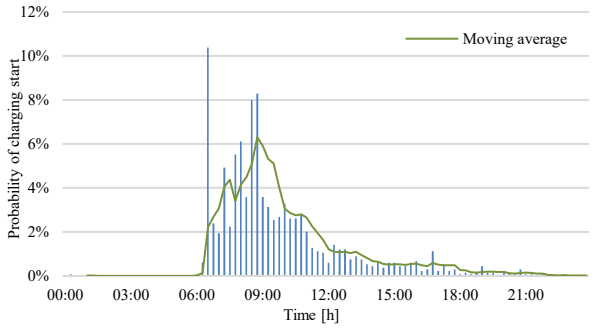


Bild 1 Wahrscheinlichkeit des Beginns des Ladevorgangs.

4.2 Analyse Ladeleistung

Die kumulierte Ladeleistung steigt morgens deutlich an, erreicht um die Mittagszeit ein Plateau und fällt am Nachmittag ab. Dieses Lastprofil folgt typischen Campus-Arbeitszeiten und lässt sich effizient mit einer typischen PV-Einspeisekurve überlagern [4]. Die meisten Fahrzeuge nutzen AC-Umrichter mit Ladeleistungen von 3,7 – 11 kW.

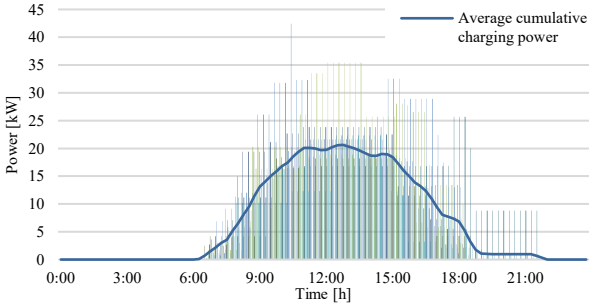


Bild 2 Ladeleistung des Ladeparks für ausgewählte repräsentative Tage.

4.3 Externe Einflussfaktoren

Wetterparameter (Temperatur, Sonnenscheindauer, Niederschlag, Wind) zeigen kein relevantes Einflussniveau (Korrelationsbereich –0.108 bis +0.053). Lediglich niedrigere Temperaturen korrelieren minimal mit erhöhter Fahrzeugzahl, jedoch ohne praktische Relevanz. Einen signifikanten Einfluss haben jedoch die Schulferien in Berlin/Brandenburg. Sie reduzieren den mittleren Energiebedarf um etwa 22 %. Wochenenden zeigen eine geringere Nutzung und spätere Startzeiten. Vorlesungszeiten haben dagegen keinen messbaren Einfluss, da die Nutzer ausschließlich Mitarbeitende sind.

Parameter	Einheit	Pearson	Spearman
Durchschnittstemperatur	[°C]	-0,099	-0,108
Sonnenscheindauer	[h]	-0,062	-0,055
Niederschlag (>1 l)	[l/m²]	-0,000	0,006
Windgeschwindigkeit	[Bft]	0,053	0,017

Tabelle 1 Korrelationskoeffizienten Energieverbrauch – meteorologische Werte

4.4 Nutzerspezifische Muster

Fünf exemplarische Nutzer zeigen unterscheidbare Muster hinsichtlich der Wahrscheinlichkeit der Anwesenheit pro Wochentag, Ladestartzeiten und Energiemengen. Die individuellen Analysen erlauben Rückschlüsse auf Fahrverhalten und Fahrzeugtypen und können zur Verbesserung probabilistischer Mobilitätsmodelle beitragen.

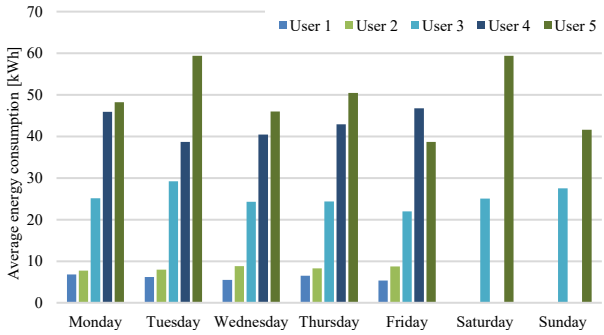


Bild 3 Durchschnittlicher Energieverbrauch pro Nutzer und Wochentag.

5 Fazit

Die Ladedaten zeigen, dass das Nutzungsverhalten nahezu vollständig durch universitäre Arbeitszeiten geprägt ist. Der Morgen bildet den dominanten Zeitraum für neue Ladevorgänge. Das Wetter spielt keine Rolle, Ferien hingegen reduzieren die Gesamtnachfrage messbar. Die resultierenden Lastprofile sind stabil, homogen und eignen sich gut als Grundlage für Lastprognosen und intelligentes Energiemanagement in ähnlichen halb-öffentlichen Ladeparks. Die Kopplung von PV-Erzeugung und typischem Ladeverlauf bietet hierbei besonderes Optimierungspotenzial.

6 Ausblick

Mit der herstellerübergreifenden Umsetzung der ISO 15118-20 wird eine deutlich detailliertere Kommunikation zwischen Fahrzeug und Ladepunkt möglich, insbesondere bzgl. State-of-Charge und Batteriemanagement [5]. Dies erlaubt präzisere Prognosen und eine dynamische Ladeplanung. Weitere Forschungspotenziale ergeben sich aus der Untersuchung temperaturabhängiger Ladeverläufe, einer PV-Lastprognose sowie der Modellierung künftiger Spitzenlasten durch Fahrzeuge mit höheren AC-Leistungen oder bidirektionalem Laden.

7 Dank

Diese Forschungsarbeiten werden im Rahmen des Verbundvorhaben EIZ: Energie-Innovationszentrum (Projektnummern 85056897 und 03SF0693A) mit Mitteln aus dem Strukturstärkungsgesetz Kohleregionen von der Bundesregierung, dem Bundesministerium für Forschung, Technologie und Raumfahrt sowie dem Land Brandenburg gefördert.

8 Literatur

[1] Q. Cui, X. Bai, S. Zhu and B. Huang, ”Cost-benefit calculation and analysis of V2G system,” 2016 China International Conference on Electricity Distribution (CICED), Xi’an, China, 2016, pp. 1-5, doi: 10.1109/CICED.2016.7576284.

[2] W. Zhang, K. Spence, R. Shao and L. Chang, ”Optimal Scheduling of Spinning Reserve and User Cost in Vehicle-to-Grid (V2G) Systems,” 2018 IEEE Energy Conversion Congress and Exposition (ECCE), Portland, OR, USA, 2018, pp. 1058-1064, doi: 10.1109/ECCE.2018.8558391.

[3] M. Caruso, P. Livreri, R. Miceli, F. Viola and M. Martino, ”EV Charging Station at University Campus,” pp. 615-623, doi: 10.1109/INTLEC.2017.8214206.

[4] S. Hovet, B. Farley, J. Perry, K. Kirsche, M. Jerue, and Z. T. H. Tse, ”Introduction of electric vehicle charging stations to university campuses: A case study for the University of Georgia from 2014 to 2017,” Batteries, vol. 4, no. 2, p. 27, 2018, doi: 10.3390/batteries4020027.

[5] DIN, “Road vehicles — Vehicle to grid communication interface — Part 20: 2nd generation network layer and application layer requirements,” DIN EN ISO 15118-20, Berlin: Beuth Verlag, 2022, doi: 10.31030/3364120.

DOI 10.5162/iCCC2026/P49

Großbrandtests mit Elektrofahrzeugen für den Transport gefährlicher Güter

Hannes Soderer, Tobias Gleim, Nils Böttcher, Rico Tschirschwitz, Frank Wille,
Bundesanstalt für Materialforschung und -prüfung (BAM), Berlin, Germany
Kontakt: hannes.soderer@bam.de

Abstract

Der zunehmende Einsatz von batterieelektrischen Transportfahrzeugen (Battery Electric Transport Vehicles – BETVs) im Verkehrswesen wirft Sicherheitsfragen beim Transport gefährlicher Güter auf. Obwohl Batteriebrände auf Zell-, Modul- und Fahrzeugebene umfassend untersucht wurden, fehlen systematische Daten zu deren thermischen Auswirkungen auf Gefahrgutverpackungen. Um diese Lücke zu schließen, führt die Bundesanstalt für Materialforschung und -prüfung (BAM) auf ihrem Testgelände Technische Sicherheit (BAM-TTS) groß angelegte Brandtests durch, in denen konservative Unfallszenarien mit unterschiedlichen Zellchemien und -kapazitäten nachgestellt werden. Die Tests kombinieren fahrzeugbedingte Brandlasten mit standardisierten Propan-Referenzbränden und erfassen mittels hochauflösender Instrumente zentrale Messgrößen wie Wärmefluss, Temperatur und Strahlung. Der daraus entstehende Datensatz liefert erstmals eine direkte Quantifizierung der thermischen Auswirkungen von BETV-Bränden auf Verpackungen und unterstützt Expertengremien bei der Bewertung, ob bestehende regulatorische Anforderungen für alternative Antriebssysteme weiterhin angemessen sind.

1 Motivation

Die sicherheitstechnische Bewertung des Transports gefährlicher Güter ist eine zentrale Aufgabe der BAM. Um diesen Auftrag zu erfüllen, betreibt die BAM umfangreiche wissenschaftliche Forschung und fungiert als verantwortliche Behörde für diesbezügliche Genehmigungsverfahren in Deutschland. Gleichzeitig bringt sich die BAM in verschiedene Gremien ein, um sicherzustellen, dass die laufende Weiterentwicklung der Vorschriften für den Transport gefährlicher Güter den aktuellen Stand der Wissenschaft und Technik reflektiert und somit die Sicherheit bei der Beförderung gewährleistet ist.

Die BAM befasst sich derzeit mit neuen Antriebstechnologien wie batterie- und wasserstoffbetriebenen Transportfahrzeugen. Es werden eine Vielzahl von Fragen, die für die Sicherheitsbewertung relevant sind, sowohl analytisch als auch experimentell untersucht. Beide Antriebstechnologien wurden in den letzten Jahren von der Industrie weiterentwickelt und befinden sich auch derzeit in einem permanenten Optimierungsprozess, sodass noch keine jahrzehntelangen Betriebserfahrungen wie beispielsweise bei Verbrennungsmotoren vorliegen. Infolgedessen sind potenzielle Unfallszenarien noch nicht vollständig identifiziert und aufgrund ihrer Komplexität schwer zu beschreiben, zu modellieren und einzuschätzen. Um diesem Problem zu begegnen, führt die BAM regelmäßig Experimente auf dem BAM-TTS durch, um realistische Unfallszenarien zu analysieren.

In einem aktuellen Forschungsprojekt wird untersucht, ob BETVs für den Transport gefährlicher Güter eingesetzt werden können und welche möglichen Einschränkungen oder Randbedingungen erfüllt sein müssen, um einen sicheren Betrieb zu gewährleisten. Für Transporte mit radioaktiven Stoffen oder explosiven Substanzen hat die BAM potenzielle Gefahrenszenarien identifiziert, in denen ein Batteriebrand die Ladung beeinträchtigen könnte.

Es gibt zahlreiche Studien zu potenziellen Batteriebränden und deren Ursachen, beispielsweise thermisches Durchgehen und mögliche Ausbreitungseffekte über verschiedene Batteriesysteme hinweg. Dazu gehören Untersuchungen zu eigenständigen Batteriesystemen, Fahrzeugbatteriebränden in offener Umgebung und Tunnelbrandszenarien. Die BAM hat jedoch eine zentrale Wissenslücke identifiziert: Derzeit gibt es keine systematischen Studien, die Batteriebrände speziell im Hinblick auf ihre Auswirkungen auf den Transport gefährlicher Güter untersuchen. Diese Lücke schränkt sowohl die Arbeit der Ausschüsse als auch die Möglichkeit umfassender behördlicher Genehmigungsverfahren ein, da die Wechselwirkung zwischen Batteriebränden und gefährlicher Fracht bisher nicht im Mittelpunkt der experimentellen Forschung stand.

2 Aktueller Stand von Wissenschaft und Technik für Batterieunfallszenarien

Der Anteil von Elektrofahrzeugen in Deutschland im Privat- und Transportverkehr steigt aufgrund politischer und ökologischer Anforderungen stetig an [1]. Infolgedessen ist die Sicherheit von batterieelektrischen Fahrzeugen (BEVs) und BETVs Gegenstand zahlreicher Untersuchungen geworden [2, 3, 4]. Um Daten über das charakteristische Verhalten dieser Fahrzeuge bei einem Brand zu sammeln, wurden verschiedene groß angelegte Experimente durchgeführt. Das Verfahren ist bei den meisten sehr ähnlich: Ein BEV-Brand wird ausgelöst, entweder durch eine externe Flammenquelle, z. B. eines Propangasbrenners, an das Fahrzeug, oder durch Beschädigung der Batterie, wodurch ein thermisches Durchgehen ausgelöst wird. Der Brand wird dann durch Berechnung der Wärmefreisetzungsrate (HRR) über die Zeit charakterisiert. Dies geschieht häufig durch Sauerstoffverbrauchskalorimetrie, vgl. [5].

Beim Vergleich der HRR von BE(T)Vs mit Verbrennungsmotorfahrzeugen (ICEVs) besteht der Hauptunterschied in lokalen Spitzenwerten (bis zu 10 MW) in der HRR für

DOI 10.5162/iCCC2026/P49

BE(T)Vs aufgrund der hohen Energiemenge, die in kurzer Zeit durch das Auftreten eines thermischen Durchgehens freigesetzt wird, vgl. [6]. Abgesehen davon unterscheidet sich die gesamte Wärmeabgabe nicht wesentlich, vgl. [7].
Weitere Phänomene, die das Handling unfallgefährdeter BE(T)Vs stark beeinflussen, sind die spontane Wiederentzündung beschädigter Batterien und die Fähigkeit von Batterien, in Abwesenheit von Sauerstoff zu brennen, da während eines Brandes chemisch gebundener Sauerstoff aus den Batteriematerialien freigesetzt wird.
Obwohl die HRR das Hauptmerkmal zur Beschreibung eines Fahrzeugbrandes ist, liefert sie keine Informationen über lokale Wärmeflüsse, z. B. in die Ladung. Das aktuelle Forschungsprojekt zielt darauf ab, eine geeignete Datenbasis zu erhalten.

3 BAM Testgelände Technische Sicherheit (BAM-TTS)

Für Tests in Realmaßstab bzw. gefährliche, explosive und lärmintensive Experimente betreibt die BAM das Testgelände Technische Sicherheit südlich von Berlin im Bundesland Brandenburg. Das BAM-TTS umfasst eine Vielzahl spezialisierter Testbereiche, darunter verschiedene Wasserstoffversuchsanlagen, ein Sprengstoffversuchsgelände, verschiedene Fallversuchs- und Brandversuchsanlagen. Diese Testanlagen ermöglichen es, ein breites Spektrum an Themen zu behandeln und wichtige Sicherheitsfragen zu beantworten.

3.1 Brandversuchsanlagen

Die BAM betreibt verschiedene Brandversuchsanlagen, die zur Untersuchung und Bewertung einer Vielzahl von Szenarien ausgelegt sind, darunter Poolfeuer, Propanbrände, Wasserstoffbrände, Holzbrände sowie Brandversuche für zerstörende und zerstörungsfreie Untersuchungen. Abbildung 1 zeigt die derzeit aktiven Versuchsanlagen auf dem Brandprüfplatz, versehen mit einer Kurzbeschreibung.
Unter den zahlreichen Brandprüfständen zeigt Abbildung 2 die zerstörungsfreie Brandprüfung mit zwei Brennerringen und einem Feuerreferenzbauteil für eine 30-minütige Brandprüfung mit Propangasflammen, vgl. [8].

3.2 Batterietestanlagen

Die Batterietestanlagen sind über mehrere Standorte auf dem BAM-Gelände verteilt. So werden beispielsweise Batterietests auf Zellebene am Hauptsitz in Berlin Steglitz durchgeführt, wo vollständige Testumgebungen und Messtechnologien zur Verfügung stehen, um Verfahren wie standardisierte Penetrationstests durchzuführen (siehe Abbildung 3). Dort werden eine Vielzahl von Batteriezellen mit unterschiedlichen Kathodenchemien untersucht und verschiedene Aspekte wie Leistung, Energiedichte, Sicherheit und Lebensdauer untersucht. Ein Team aus dem Tätigkeitsbereich Elektrische Energiespeicherung und -umwandlung analysiert den Sicherheitszustand (State-of-Safety) von Batterien.

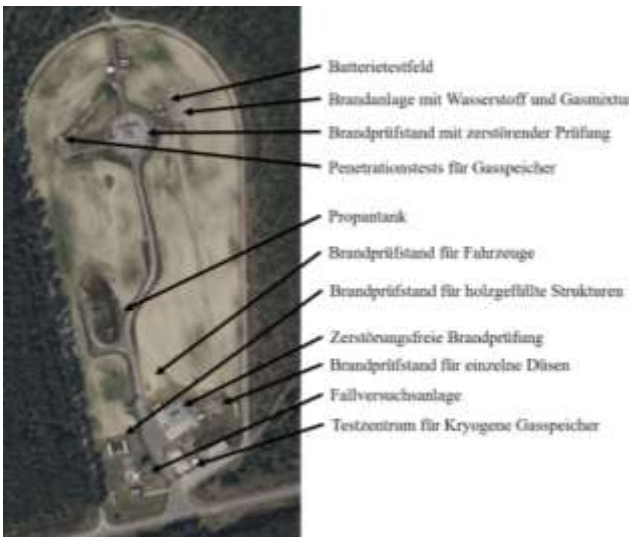


Abbildung 1 Brandprüfplatz mit momentan aktiven Versuchsanlagen

Abbildung 2 Feuerreferenzbehälter in einem Brandprüfstand mit zwei Brennerringen, vgl. [8].

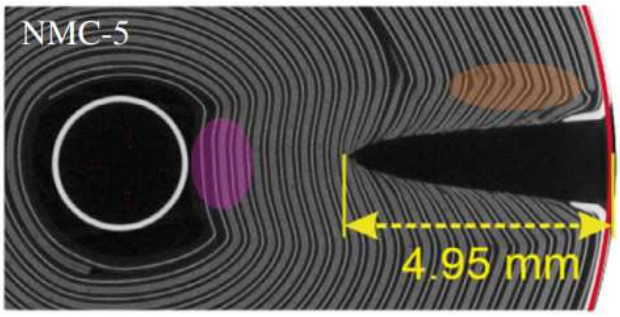


Abbildung 3 Hochpräzises Nageltest Setup zum kontrollierten Erzwingen des thermischen Durchgehens, vgl. [10].

Untersuchungen von großen Batteriezellen [11], Batteriemodulen [12] und großen Batteriesystemen werden auf dem BAM-TTS auf einer speziellen Plattform durchgeführt, die es ermöglicht, eine Vielzahl von Batterietypen unter kontrollierten Bedingungen zu untersuchen. Diese Anordnung (vgl. Abbildung 4) ermöglicht es, das Verhalten von Batterien in verschiedenen Betriebs- und Fehlerszenarien zu untersuchen, wie z. B. Laden und Entladen, Überladen (vgl. Abbildung 5) oder interne Ausfälle. Darüber hinaus wurde bereits das Ausbreitungsverhalten von thermischen Durchgangsreaktionen zwischen Modulen untersucht, um verschiedene Materialien als Barriere gegen

DOI 10.5162/iCCC2026/P49

thermisches Durchgehen zu qualifizieren. Der verwendete Prüfstand ist bereits mit einem Hochdruck-Wassernebel-Löschsystem ausgestattet. Dies ermöglicht Tests zum thermischen Durchgehen mit Löschintervention unter dem Gesichtspunkt eines sparsamen Wasserverbrauchs.

Abbildung 4 Batterieprüfstand für Module und reale Batterien mit 1) Module ausgestattet mit Thermoelementen, 2) Versuchstisch, 3) - 4) Diodenlaser-Spektrometer für HF-Konzentration (Flusssäure), 5) Nahfeld-Thermoelemente, 6) Abzugshaube, 7) Hochdruck-Wassernebelanlage und 8) Kamerasystem, vgl. [9, 16].

Ein zentraler Schwerpunkt ist die Lebensdauer der Batterien, insbesondere im Hinblick auf Second-Life-Anwendungen, bei denen Batterien aus BEVs als stationäre Energiespeichersysteme wiederverwendet werden. Das BAM-TTS ermöglicht realistische Testbedingungen zur Bewertung sicherheitskritischer Faktoren wie thermische Instabilität, Gasfreisetzung und potenzielle Explosionsrisiken (vgl. Abbildung 6), insbesondere bei alternden Zellen, die für den weiteren Gebrauch vorgesehen sind.

Abbildung 5 Thermisches Durchgehen durch Überladen eines 6,85kWh-Moduls [13].

4 Versuchsaufbau auf den Brandversuchsanlagen

Der Versuchsaufbau basiert auf konservativen, praxisorientierten Unfallszenarien, die potenzielle Brandereignisse mit BETVs darstellen, die sich auf ein Paket oder eine Ladung mit gefährlichen Gütern auswirken können (siehe Abbildung 7). Diese Szenarien werden aus Analysen möglicher Schadensverläufe abgeleitet, wobei die Art und Kapazität der Batterien (z. B. NMC, LFP, NCA/LTO), ihre Einbaulage im Fahrzeug, zusätzliche brennbare Komponenten, Umgebungsbedingungen wie Tunnel, Brücken oder offene Flächen sowie das zeitabhängige Brandverhalten einschließlich Brandausbreitung und potenzieller Wiederentzündung berücksichtigt werden. Die resultierenden Szenarien werden so ausgewählt, dass sie sowohl realistisch als auch ausreichend konservativ sind, um behördliche Bewertungen gemäß den relevanten Normen zu unterstützen.

Abbildung 6 Gasexplosion infolge des Ausgasens verursacht durch thermisches Durchgehen eines 2,65kWh-Moduls [14].

Für fahrzeugbezogene Brandversuche, wie in Abbildung 7 dargestellt, ist das TTS, insbesondere der in Abbildung 1 dargestellte Brandprüfplatz, ideal geeignet. In Kombination mit dem Batterieteststand für stationäre Batterien (siehe Abbildung 4) werden verschiedene Unfallszenarien mit Gefahrstoffen analysiert, bewertet, definiert und anschließend in einer realistischen Testumgebung untersucht, beispielsweise unter Verwendung eines BETV und einer Transportverpackung. Ein spezielles Feuerreferenzpaket, wie in Abbildung 2 dargestellt, wird verwendet, um Experten aus verschiedenen Disziplinen standardisierte Messdaten zur Verfügung zu stellen.

Abbildung 8 Schematischer experimenteller Aufbau mit BETV, Feuerreferenzbehälter [18, 19], und Batterie als Brandlast.

Auf der Grundlage dieser Randbedingungen werden unter Verwendung der vorhandenen Brandprüfstände und der kalorimetrischen Infrastruktur auf dem BAM-TTS realmaßstäbliche Versuchsaufbauten konzipiert. Der Versuchsaufbau umfasst definierte Brandquellen sowie zusätzliche Brandlasten aus BETV-Komponenten wie Holz, Reifen, Öl, Kunststoffen oder Zusatzbatterien, um repräsentative thermische Bedingungen zu reproduzieren. Feuerreferenzbehälter werden innerhalb des Versuchsaufbaus positioniert, um standardisierte Messungen der Wärmezufuhr und der thermischen Auswirkungen zu ermöglichen, vgl. [15].

Umfassende Brandversuche werden unter kontrollierten und vollständig dokumentierten Bedingungen durchgeführt, wobei die drei Batterietypen systematisch über verschiedene Kapazitätsbereiche (z. B. 90 - 220 kWh) variiert werden. Während jedes Experiments werden wichtige physikalische Größen wie Temperaturverteilungen, Wärmefluss und Wärmestrahlung aufgezeichnet. Hochauflösende Messgeräte, unterstützt durch Video- und Wärmebildaufnahmen, gewährleisten eine umfassende und zuverlässige Datenerfassung.

5 Zusammenfassung und Ausblick

Der zunehmende Einsatz von BETVs für den Transport gefährlicher Güter bringt neue Herausforderungen mit sich, denen die aktuellen Sicherheitsanforderungen und Prüfnormen noch nicht ausreichend Rechnung tragen. So müssen beispielsweise die bestehenden thermischen Anforderungen der IAEO, wie die 30-minütige Brandprüfung bei 800°C [17] und die damit verbundenen Bewertungskriterien, unter Berücksichtigung der spezifischen Risiken, die mit diesen neuen Antriebstechnologien verbunden sind, detailliert überprüft werden.
Derzeit gibt es keine experimentellen Studien, die die Wärmeflüsse auf Verpackungen während Bränden mit Fahrzeugen mit alternativen Antrieben systematisch quantifizieren. Die verfügbaren Daten stammen fast ausschließlich aus Batterietests auf Zell- oder Modulebene und spiegeln nicht die realen Transportszenarien wider, in denen die thermischen Auswirkungen auf eine Verpackung oder deren Inhalt von zentraler Bedeutung sind. Bestehende groß angelegte BEV-Brandtests befassen sich mit anderen Fragestellungen und können nicht auf BETVs für den Transport gefährlicher Güter übertragen werden.
Das geplante Forschungsprojekt zielt darauf ab, diese Lücke durch realmaßstäbliche Brandversuche mit BETVs im Gefahrengutkontext zu schließen, unterstützt durch numerische Analysen zur Darstellung realistischer thermischer Belastungsszenarien. Es werden verschiedene Hochvolt-Batterietypen mit unterschiedlichen Kapazitäten und Tellchemien (z. B. NMC, LFP, NCA/LTO) untersucht, um technologiespezifische Unterschiede und Skalierungseffekte zu erfassen. Die Ergebnisse werden den verschiedenen Ausschüssen und Gremien helfen, die Auswirkungen der Wärmezufuhr im Zusammenhang mit der Ladung zu bewerten und gegebenenfalls geeignete Maßnahmen einzuleiten.

6 Literatur

[1] KBA: Neuzulassungen von Kraftfahrzeugen mit alternativem Antrieb im Oktober 2025: https://www.kba.de/DE/Statistik/Produktkatalog/produkte/Fahrzeuge/fz28/fz28_gentab.html

[2] M. Neske et. al.: Evaluierung von technischen Verfahren zur Löschmitteleinbringung in Hochvoltspeicher, Institut für Brand- und Katastrophenschutz Heyrothsberge Abteilung Forschung - Institut der Feuerwehr–, Forschungsbericht, 2025

[3] R. Tschirschwitz et. al.: BEV Accidents-Consequences in Case of a Battery Thermal Runaway, Proceedings of the 31st Annual Congress of the European Association for Accident Research and Analysis (EVU), 2023

[4] A. Lecocq et. al.: Comparison of the fire consequences of an electric vehicle and an internal combustion engine vehicle, 2. International Conference on Fire in Vehicles - FIVE 2012, Chicago 2012

[5] C. Huggett: Estimation of Rate of Heat Release by Means of Oxygen Consumption Measurements, Fire and Materials, 1980

[6] M. Siemon, M. Rupp: Brandszenarien bei E-PKW – Stand der Wissenschaft, KSI Brandschutz AG, 2022

[7] P.-J. Sturm et al.: BRAFA – Brandauswirkungen von Fahrzeugen mit alternativen Antriebssystemen, TU Graz, 2021

[8] T. Quercetti et. al.: A FIRE TEST STAND FOR THERMAL TESTING OF EXTRA-LARGE PACKAGES, Proceedings of the ASME 2025 Pressure Vessels & Piping® Conference, 2025.

[9] SEE 2L Project (BAM, OVGU & vfdb), Safety of Electrochemical Energy Storage in Second Life Applications, 2023

[10] N. Böttcher et. al.: High Precision Nail-Penetration Setup for the Controlled Thermal Runaway Initiation of Lithium-Ion Cells at Very Low Temperatures, Energy Technology, DOI: 10.1002/ente.20230137, 2024

[11] T. Rappsilber et. al.: Hazard classification of different lithium-ion battery types after triggering thermal runaway, Journal of Loss Prevention in the Process Industries, 2025

[12] R. Tschirschwitz et. al.: Harmful effects of lithium-ion battery thermal runaway: scale-up tests from cell to second-life modules, RSC advances, 2023

[13] https://www.youtube.com/watch?v=0hIzS-0Ajgo

[14] https://www.youtube.com/watch?v=QwXT TxLQ9mA

[15] M. Feldkamp, T. Gleim, F. Wille: Evaluation of heat fluxes in Fire Reference Test conducted in BAM propane gas Fire Test Facility, presented at the 27th International Conference on Structural Mechanics in Reactor Technology, Yokohama, Japan, 04.03.2024, 2024.

[16] S.-K. Hahn et. al.: Forschung zur Sicherheit von Elektroenergiespeichern, Institut für Schadenverhütung und Schadenforschung der öffentlichen Versicherer e.V. (IFS), 2025

[17] IAEA: Regulations for the Safe Transport of Radioactive Material - Specific Safety Requirements No. SSR-6 (Rev.1), IAEA, 2018

[18] M. Feldkamp, T. Quercetti, F. Wille: Outcomes of three large scale fire reference tests conducted in BAM fire test facility, Proceedings of the ASME 2020 Pressure Vessels & Piping Conference, 2020.

[19] M. Feldkamp et. al.: R&D Activities by BAM Related to Transport Package Fire Testing, Proceedings of the 20th International Symposium on the Packaging and Transportation of Radioactive Materials, 2022.

DOI 10.5162/iCCC2026/P50

Drone Usage in industrial fields

Phillip Werban, Matthias Nattke and Peggy Naeser
BTU Cottbus-Senftenberg, Cottbus, Germany
Phillip.Werban#1@b-tu.de

Abstract

Drones are increasingly used across industrial environments for monitoring, inspection, and safety-related tasks. While their ability to cover large areas and collect high-quality data provides significant operational advantages, most industrial deployments still rely on manual piloting, limiting scalability and efficiency. This paper examines the distinction between autonomous, semi-autonomous, and non-autonomous drone operations and highlights the benefits of integrating drones into industrial software systems such as maintenance, security, and monitoring platforms. By analyzing typical use cases, including solar field inspection, surveillance of industrial sites, and early fire detection in open-cast mining areas, this work discusses when autonomous and integrated systems offer measurable advantages and when manual operation remains more suitable. In addition, regulatory constraints for unmanned aerial vehicle (UAV) operations are considered, with a focus on the risk-based classification into open, specific and certified categories. The findings show that recurring and predictable missions benefit most from autonomous and integrated approaches, while irregular or highly dynamic scenarios still require human intervention. Due to the frequent use of Beyond Visual Line of Sight (BVLOS) operations in industrial contexts, autonomous drone missions typically fall into the specific category, requiring structured risk assessment and operational documentation. Overall, integrated autonomous drones can significantly reduce labor effort, increase data availability, enable compliant BVLOS operations and improve decision-making in industrial workflows.

1 Introduction

In recent years, drones have become omnipresent in public awareness and in the media. The industry has also shown growing interest in unmanned aerial systems. The inspection of large solar parks, monitoring of extensive industrial machinery and other applications show the significant potential UAVs can offer for improving efficiency. However, the industry rely still for many scenarios on a human pilot, which makes these operations costly and labor-intensive. Especially in recurring missions, autonomous flights could provide many benefits like reproducible procedures, reduced operating expenses through a decreased need for specialized personnel [1].

2 Industrial Use of Drones

Typical industrial applications of drones span a broad spectrum of monitoring, inspection and surveillance tasks [2]. For instance, drones are increasingly used to monitor the condition and performance of large solar fields. Drones can cover extensive areas quickly and with help of a thermal camera identify defects in PV-Installations like for example hotspots, outages and Strings [3]. They can also have an important role in the surveillance of construction sites and company sites as support for security personnel by providing aerial perspectives that enable more effective detection of unauthorized access, hazardous situations and criminal actions. For example a combination of Camera surveillance, motion detectors and a drone with a drone-dock can be used to protect construction sites from material theft. Another use-case for drones can be as an early fire detection in open-cast coal mining areas. Their drones can be equipped with thermal sensors to identify potential ignition points and embers to report potential fires before they become critical [4].

Despite their suitability for recurring tasks in large areas, most drone missions today are still performed under direct control of a human pilot, which limits the level of possible automation and scalability in industrial workflows.

3 Autonomous and Integrated Drones

Drones can be called autonomous, when they are capable of executing missions without direct human control. They rely on pre-programmed flight paths, sensors and mission-based decision-making to navigate and complete their missions. That includes also the deployment of the drone itself. This is mostly accomplished by the use of drone docks. The functionality for the docks include maintaining a connection to the internet, charging and housing the drone while not on a mission, often collecting weather data for pre-flight checks and sometimes providing processing capabilities for the missions.
A drone can also operate semi-autonomous, when a mission is pre-planned and deployed directly on the drone. The pilot then only has to start the mission and monitor the flight.
The integration of a drone describes a drone system that is connected to broader industrial systems like maintenance management, security systems and monitoring platforms [5]. This enables for example an automatic data collection with processing and possible triggering of follow-up actions like calling the fire department when a fire is detected. Two Paths can be identified for an integrated drone, time critical events and data analysis. The detection of a fire is for example a time critical event, where the collected data is less important than the alarm of the fire department and the control Center. In case of data analysis, the quality of the collected data is more

DOI 10.5162/iCCC2026/P50

important than a quick notification. A daily cartography of a construction site can be an example for this.
In **Figure 1** a possible path from the drone collected data to their endpoints in the control Center as well into the industrial systems is shown for an integrated drone. In contrast **Figure 2** shows the path for the data, when the drone is not integrated. In this case there is a time delay between the collecting of the data and upload to the processing server over the computer of the drone pilot. Because of this delay, the processed data is not current and therefore not needed in the control Center [6].
An autonomous and integrated drone is both able to execute missions without direct human control and is connected to the industrial systems to use, process and react to the collected data. This offers clear advantages in repetitive and standardized missions such as the routine inspection of solar fields or industrial machinery where it reduces labor costs, improves efficiency by creating maintenance plans and ensures consistent results through the repetitiveness.

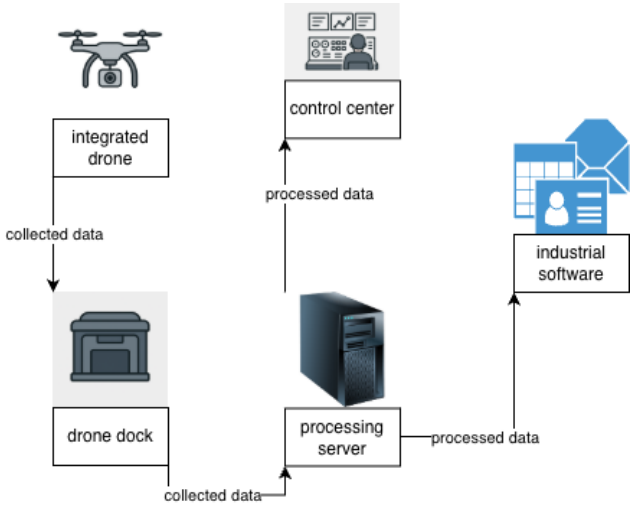


Figure 1 Shows the path of the data from collection on the drone to their endpoints in the industrial systems and the control Center for an integrated drone.

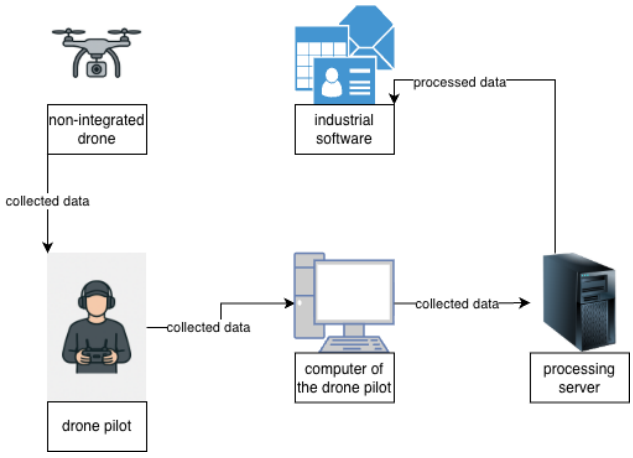


Figure 2 Shows the path of the data from collection on the drone to their endpoints in the industrial systems and the control Center for a non-integrated drone.

4 Autonomous BVLOS Drone Operations and Regulatory Classification

UAV operations are regulated using a risk-based classification framework, which groups missions into different operational categories depending on their complexity and associated safety risks. This framework defines the applicable technical, operational and organizational requirements for drone operations and ensures a proportionate level of regulatory oversight. The three categories: open, specific and certified differ significantly in terms of operational limitations, approval requirements and certification obligations. These categories are also presented in Figure 3.

Figure 3 Overview of the regulatory categories for UAV operations (open, specific and certified) and their key operational criteria. [7]

4.1 Open Category

The open category is the least restrictive regulatory category for UAV operations and is intended for low-risk scenarios. Operations in this category must meet all defined criteria, including a maximum take-off mass of less than 25kg, operation below 120m above ground level (AGL) and flights conducted exclusively in Visual Line of Sight (VLOS).
Furthermore, flights in the open category are not permitted over assemblies of people and must not involve the transport of dangerous goods. Due to these strict limitations, operations in the open category are restricted to simple mission profiles with low operational complexity and do not require prior authorization from aviation authorities.

4.2 Special Category

The specific category covers UAV operations that exceed at least one limitation of the open category. This includes operations with an aircraft mass greater than 25 kg, flights above 120m AGL or within special or controlled airspace, as well as Beyond Visual Line of Sight (BVLOS) missions.
Because these operations introduce increased operational complexity and risk, they cannot be conducted under the open category framework. Instead operations in the specific category require a structured risk assessment as

DOI 10.5162/iCCC2026/P50

an operational manual and formal authorization by the competent authority, allowing more advanced and flexible mission profiles while maintaining an appropriate level of safety.

4.3 Certified Category

The certified category applies to UAV operations that meet at least one high-risk criterion, such as flights over assemblies of people, the transport of dangerous goods or the transport of people. These operations are considered comparable to manned aviation in terms of safety impact and potential consequences.
As a result, the certified category is subject to the most stringent regulatory requirements, including aircraft certification, operator certification and, where applicable, remote pilot licensing. Due to the extensive certification effort involved, this category is reserved for highly safety-critical operations and is not typically used for conventional BVLOS drone missions.

4.4 Operational Manual

For operations conducted outside the open category, particularly within the specific category, the development of a comprehensive operations manual is mandatory. The operations manual serves as a central document describing the concept of operations, operational procedures, system architecture, safety measures, emergency procedures and the allocation of responsibilities.
This document is a key requirement for regulatory approval and demonstrates that identified risks are adequately mitigated. In the context of BVLOS operations, the operations manual plays a crucial role in ensuring safe, standardized and compliant mission execution.

4.5 Autonomous BVLOS Drone Operations

Autonomous integrated drones are operated in Beyond Visual Line of Sight (BVLOS) missions, meaning that the remote pilot is unable to maintain direct visual contact with the aircraft at all times. BVLOS operations typically rely on technical systems such as sensors, communication links and automation to ensure safe navigation, separation from other airspace users and mission execution.
Due to the BVLOS nature of autonomous drone operations, compliance with the open category is not possible. Consequently, these missions must be conducted under the specific category, which mandates the development of a comprehensive operations manual.

5 Integrated Autonomous Flights vs. Semi-Autonomous Flights vs. non-Autonomous Flights

Industrial applications can be sorted into two categories: recurring scenarios and one or few time scenarios.
When a task for a drone is a one or few time event, the costs for integration are greater than the advantages afterwards. In this case, a drone can be used semi-autonomous when the flight path of the mission is predictable before the flight, so that the drone can be deployed by the pilot and then fly the mission on its own. If the flightpath of the mission is unclear or the mission needs a human pilot, then the non-autonomous Flight is the only option.

6 Conclusion

Drones can be useful in many industrial applications, ranging from monitoring and surveillance to data collection. The degree of automation for the drone missions depends mainly on the mission itself. If the task is recurring and predictable, a suitable level of automation can be achieved through autonomous drone flights based on pre-planned missions. The data collected by the drone can be used for direct reactions, such as intruder detection for perimeter security or automatically alerting the fire department in the event of fire detection. Integrating the drone into industrial systems can provide significant benefits. These benefits include a reduction in specialized personnel requirements, lower labor costs and highly reproducible operational procedures.
In addition to technical and organizational aspects, regulatory constraints play a decisive role in the design and deployment of autonomous drone systems. UAV operations are governed by a risk-based classification framework comprising the open, specific and certified categories, each imposing different operational limitations and approval requirements. Due to the use of BVLOS operations, autonomous integrated drone missions cannot be conducted within the open category and must therefore be assigned to the specific category.
As a consequence, such operations require a structured risk assessment and the preparation of a comprehensive operations manual. This documentation forms the basis for regulatory approval and ensures that operational risks are adequately addressed, particularly in complex BVLOS scenarios. The regulatory framework therefore directly influences both the technical design and the operational integration of autonomous drone systems.
The two paths of an integrated drone system remain an important design consideration. When a mission is based on detecting time-critical events, direct integration of the drone into operational systems is essential to prevent delays. The data analysis path, on the other hand, is most suitable when the quality of processing and analysis is more important than immediate response time.

7 Acknowledgement

The Project EDISON is co-financed by the European Regional Development Fund (EFRE) on the basis of the directive issued by the Ministry of Science, Research and Culture (MWFK) to promote the strengthening of technological and application-oriented research in research networks of scientific institutions in the state of Brandenburg (StaF network) and the state of Brandenburg. The authors are responsible for the content of this publication.

DOI 10.5162/iCCC2026/P50

8 Reference

[1] L. M. Horton, K. M. Cabral, B. W. Surgenor, S. N. Givigi, and J. E. Woods, “A Framework for Autonomous Inspection of Bridge Infrastructure using Uncrewed Aerial Vehicles”, in 2024 IEEE International Systemes Conference (SysCon), pp. 1-8, IEEE, 2024

[2] M. Javaid, I. H. Khan, R. P. Singh, S. Rab, and R. Suman, “Exploring contributions of drones towards industry 4.0,” Industrial Robot: the international journal of robotics research and application, vol. 49, no. 3, pp. 476–490, 2022.

[3] Y. Zhang, F. Zhang, B. Zhang, J. Mo, Y. Xie, and Y. Qiu, “Research of uav inspection system for new energy power stations in special scenarios,” in 2025 4th International Conference on Smart Grid and Green Energy (IC-SGGE), pp. 127–130, IEEE, 2025.

[4] Z. Shao, Y. Li, R. Deng, D. Wang, and X. Zhong, “Three-dimensional-imaging thermal surfaces of coal fires based on uav thermal infrared data,” International Journal of Remote Sensing, vol. 42, no. 2, pp. 672–692, 2021.

[5] Y. Yigit, L. D. Nguyen, M. Ozdem, O. K. Kinaci, T. Hoang, B. Canberk, and T. Q. Duong, “Twinport: 5g drone-assisted data collection with digital twin for smart seaports,” Scientific reports, vol. 13, no. 1, p. 12310, 2023.

[6] G. Aiello, F. Hopps, D. Santisi, and M. Venticinque, “The employment of unmanned aerial vehicles for analyzing and mitigating disaster risks in industrial sites,” IEEE transactions on engineering management, vol. 67, no. 3, pp. 519–530, 2020.

[7] https://lba-openuav.de/en/air-law/uas-classification/, Last visited: 27.01.2026

DOI 10.5162/iCCC2026/P51

Accelerating the development of electric propulsion systems for aviation through building a digital ecosystem

Magdalena Peciak, CHESCO GmbH, Germany, magdalena.peciak@chesco.de
Heiko Witte, CHESCO GmbH, Cottbus, Germany, heiko.witte@chesco.de
Jörg Engelmann, CHESCO GmbH, Cottbus, Germany, joerg.engelmann@chesco.de

Kurzfassung

Eine signifikante Reduzierung der Treibhausgasemissionen im Luftverkehr kann nur durch den Einsatz alternativer Antriebssysteme erreicht werden. Ihre Anwendung erfordert eine Analyse des gesamten Ökosystems, einschließlich der Luftverkehrsinfrastruktur, Sicherheitsaspekte, Zertifizierung und Wartung von Flugzeugen mit neuen Antriebssystemen. Im Rahmen des DIREKT-Projekts entwickelt CHESCO einen digitalen Zwilling des (hybrid-)elektrischen Flugzeug-Ökosystems, der die Analyse verschiedener Aspekte im Zusammenhang mit alternativen Antriebssystemen ermöglicht. Derzeit werden innerhalb dieses Ökosystems drei Modelle entwickelt: Betriebs-, Batterie- und alternatives Antriebssystem. Das erste Modell analysiert die Auswirkungen des neuen Antriebssystems auf den Flugbetrieb, während das Batteriemodell die Simulation und Analyse von Änderungen der Batterieparameter während des Einsatzes ermöglicht. Das Modell für alternative Antriebssysteme konzentriert sich auf die Analyse der erforderlichen Betriebsparameter eines Flugzeugs mit einem neuen Antriebssystem. Die Verwendung eines digitalen Zwillings des (hybrid-)elektrischen Flugzeug-Ökosystems ermöglicht eine eingehende Analyse verschiedener Aspekte der Anwendung des neuen Antriebssystems sowie eine Beschleunigung seines Entwicklungsprozesses, wodurch die Projektkosten gesenkt werden.

Abstract

A significant reduction in greenhouse gas emissions in aviation can only be achieved through the use of alternative propulsion systems. Their application requires an analysis of the entire ecosystem, including aviation infrastructure, safety aspects, certification and maintenance of aircraft with new propulsion systems. In the DIREKT project, CHESCO is developing a Digital Twin of the (hybrid)-electric aircraft ecosystem that allows for the analysis of various aspects related to alternative propulsion systems. Three models are currently being developed within this ecosystem: operational, battery and alternative propulsion system. The first model analyses the impact of the new propulsion system on flight operations, while the battery model allows for simulation and analysis of changes in battery parameters during use. The alternative propulsion system model focuses on analysing the required operating parameters of an aircraft with a new propulsion system. The use of a Digital Twin of the (hybrid)-electric aircraft ecosystem allows for an in-depth analysis of various aspects of the new propulsion system's application, as well as accelerating its development process, thereby reducing project costs.

1 Introduction

Aviation is responsible for approximately 3% of global CO_2 emissions, which are the only ones that occur at altitudes. Due to the high energy demand needed for aircraft to fly and the predicted increase in the number of flight operations in the coming years, the decarbonisation of aviation is crucial to halting climate change [1].

The greatest impact on reducing greenhouse gas emissions from aviation would be achieved using new, alternative propulsion systems: all-electric, hybrid-electric and electric based on hydrogen fuel cells. However, their use requires a broader view of the entire aviation ecosystem: adaptation of airport infrastructure, safety, ensuring sufficient electricity for charging, changes in the design and maintenance of aircraft with such propulsion systems, and battery recycling.

An additional aspect that should be considered when designing a sustainable aviation ecosystem is the current level of development and pace of advancement of subsystems that make up alternative propulsion systems. The developing ecosystem must be able to consider the changing parameters of these subsystems. The current level of technology allows for the use of alternative propulsion systems in smaller aircraft flying on short regional routes: electric in GA aircraft and hybrid -electric in the segment of smaller commercial aircraft, e.g. the DHC-8 family [2].

At CHESCO, the foundations for research and development of advanced electric and hybrid propulsion systems for aviation are created. The setup of capabilities – from design through manufacturing towards test and verification are key for achieving this. At same time a thorough understanding of all stakeholders and boundary conditions is important – data from all aspects of the product lifecycle are required to understand the to be built ecosystem (Fig. 1).

CHESCO focuses on building the manufacturing and testing capabilities, with the digital process chains connecting the elements of the product lifecycle. Through understanding of the aviation ecosystem, through a thorough stakeholder analysis - and through gathering of actual data – and generic data within the aerospace landscape we are building an ecosystem which will help us and our partners to understand, what technologies are required, how to achieve

DOI 10.5162/iCCC2026/P51

airworthiness for sustainable propulsion systems and how a business model could be achieved around the various aspects of electrified aviation.

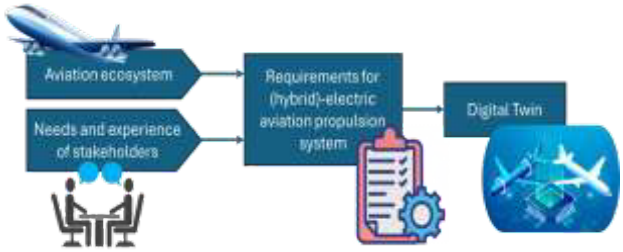


Figure 1. CHESCO's approach to creating a digital twin

2 Sustainable aviation ecosystem

The use of alternative propulsion systems in aviation involves several changes to the existing ecosystem. These adjustments are related to the emergence of new airworthiness requirements, stakeholders' needs and propulsion capabilities. One of the main changes is the adaptation of the airport infrastructure to the requirements of alternative propulsion aircraft: it needs to ensure that aircraft batteries can be recharged, required procedures need to be developed accordingly, from where the electricity will be drawn and how it will be produced. The performance of alternatively propelled aircraft is more limited and fewer adapted airports to (hybrid)-electric propulsion introduce a new level of complexity in the use of such aeroplanes. Access to aerodromes that allow full service of alternatively-powered aircraft influences the planning of flight operations of these aircraft, e.g. by scheduling longer layovers at such an aerodrome to recharge the batteries to the required level for subsequent flights from an airport that does not have this capability. Knowing the parameters of the planned flight operations also allows a tailor-made approach to alternative propulsion design in terms of providing the required range and power by selecting the appropriate battery parameters and thus adapting the airframe to it.

Part of the LuFo funded research project DIREKT (Digitaler Lebenszyklus hybrid-elektrischer Antriebssysteme) is the development of Digital Twin of the (hybrid)-electric aircraft ecosystem with using Model-Based Systems Engineering (MBSE) methodology to develop the relevant models describing different aspects of the alternatively-powered aeroplane. Currently, the following models are being developed: operational, battery and alternative propulsion systems.

2.1 Operational model

The operational model (Fig. 2) will simulate the flight of a (hybrid) electric aircraft, considering the main parameters of the ecosystem. The model will allow the simulation and analysis of various operational scenarios, including the impact of alternative propulsion on planning longer layovers at airports with the necessary battery charging infrastructure and the analysis of maximum passenger numbers for aircraft upgraded to the new propulsion system. In addition, the model will have the ability to compare different battery types for the aircraft under analysis and its operations. It will also measure the aircraft's total CO_2 emissions in relation to a defined energy mix for a region or country needed to generate the required amount of electricity to charge the aircraft's batteries. The propulsion system model will be used to analyse the required propulsion parameters, to simulate the interactions between systems and subsystems and to assist manufacturers in designing aircraft with the new propulsion system. The model will also enable the selection of an appropriate operating strategy for the hybrid-electric propulsion system and its optimisation. An in-depth analysis of the propulsion system at an early design stage will reduce errors in the next phases of aircraft development.

Figure 2. Overview of the operational model parameters: Inputs and outputs

This model also allows for a trade study to be conducted regarding the degree of hybridisation of the propulsion system of a selected aircraft (Fig. 3).

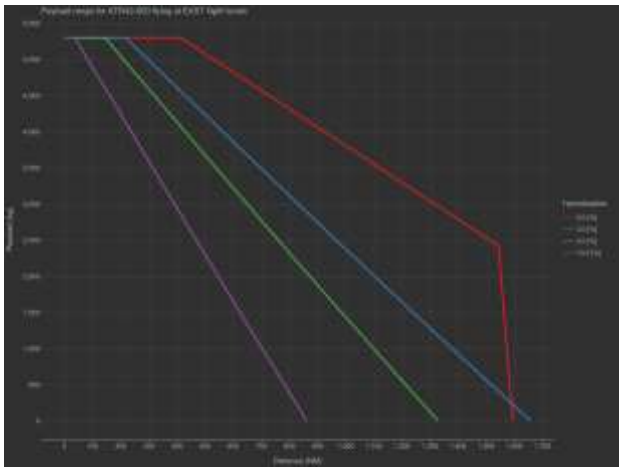


Figure 3. Payload range analysis for ATR 42-600 with different degree of hybridisation: red line- 0%, blue line- 3%, green line- 5%, purple line- 10%

2.2 Battery model

The battery model is developed using the built-in MATLAB application 'Battery Builder'. This application allows for individual design of battery packs based on actual cells from various manufacturers or cells with user-defined parameters. The battery model also simulates the thermal behaviour of such a battery pack. Figure 4 shows the battery layout, and Table 1 shows its main parameters.

DOI 10.5162/iCCC2026/P51

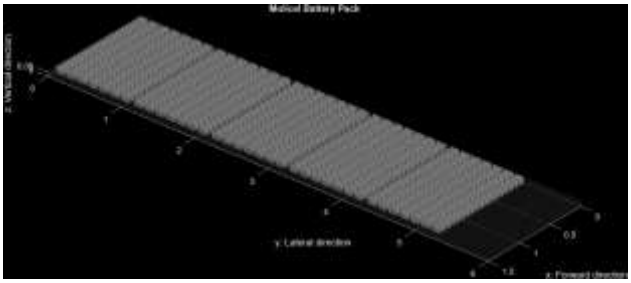


Figure 4. Molicel battery pack designed in Battery Builder App for battery simulation model

Battery parameters	
Cell	Molicel INR21700-P45B
Capacity	252 Ah
Energy	113.4 kWh
Weight	490 kg
Configuration	125s56p

Table 1. Battery parameters

The simulation model (Fig. 5) of the battery analyses its performance during operation (discharging) and charging, number of possible operating cycles, and battery State of Health (SOH). The function of this model is not only comprehensive battery analysis but also support in the certification process and reduction of the number of physical tests, which contributes to lower project costs.

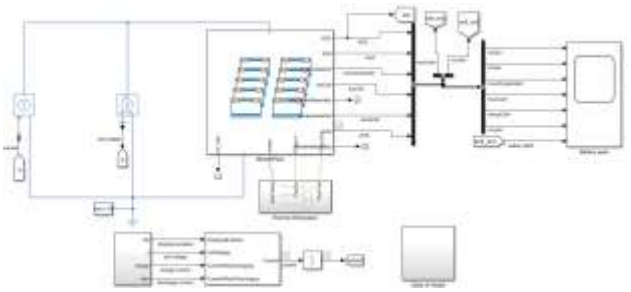


Figure 5. Battery model developed in MATLAB/Simulink software

The battery simulation model allows for the simulation of various battery charging and discharging scenarios, enabling the creation of a diverse flight schedule for the aircraft. With the help of Stateflow, which enables the creation of control logic for a simulation model, a scenario was developed that allows battery charging to be defined in various ways. The following scenarios were developed:

- charging the battery to a specific level (e.g. to 97%),
- charging the battery for a specific period of time (e.g. 30 minutes).

The first is used to determine the time required between flights to reach the required battery charge level. The purpose of the second charging scenario is to check the level to which the battery will be charged during limited downtime between flights. In both scenarios, the batteries are charged using the CC-CV (Constant Current-Constant Voltage) method.

Battery charging and discharging is characterised by electric current, which is defined by the user. This makes it possible to simulate chargers with different power ratings. For example, a flight mission was created for a Tecnam P2006T aircraft virtually retrofitted with an electric propulsion system, which included three flights of 100 km each. Between the first and second flights, the aircraft's batteries were charged to 97% (charging current 200 A, charging time 46 minutes). For the second charge, the aircraft was charged for 30 minutes on a charger with a charging current of 250 A. As a result, it reached 83.10% battery charge level. The simulation results are shown in Fig. 6 and 7.

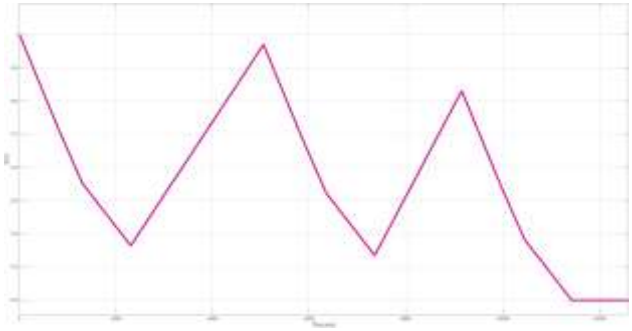


Figure 6. Battery State of Charge during the simulated flight mission

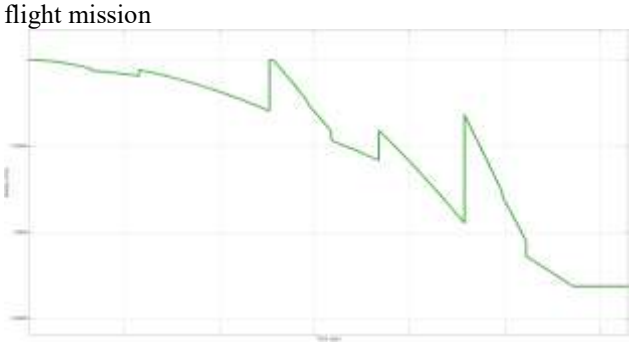


Figure 7. Battery State of Health during the simulated flight mission

2.3 Alternative propulsion system model

The above models are supplemented by an alternative propulsion system model (Fig. 8) developed in MATLAB/Simulink software, which allows for the estimation of the required operating parameters of the main subsystems of the propulsion system. This model was used to select the battery parameters for the electric Tecnam P2006T, the simulation results of which are presented in section 2.2.

The developed model allows for simulations of four types of alternative propulsion systems:

- all-electric (battery only)
- series hybrid-electric
- parallel hybrid-electric
- electric based on hydrogen fuel cells.

The application of this model is limited only to GA aircraft.

DOI 10.5162/iCCC2026/P51

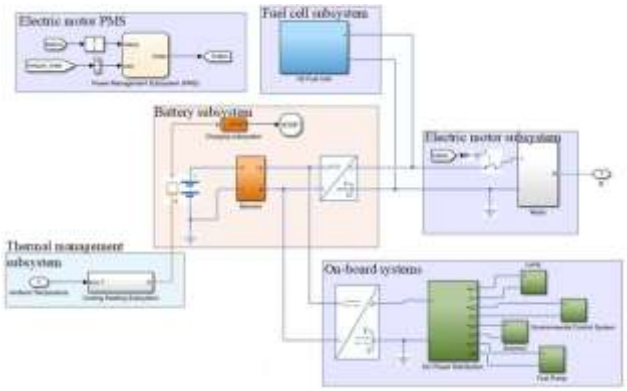


Figure 8. Alternative propulsion system model

The simulation model of alternative propulsion system (Fig. 9) consists of four main modules: flight control module, environment, airframe and propulsion. The flight control module plays the role of the pilot in the model and manages the flight mission parameters during the simulation. For each type of alternative propulsion system, the flight control module is adjusted accordingly, due to the different or additional parameters that characterise the propulsion system. Parameters from the flight control module are passed to the environment and airframe modules. The environment module is used to calculate the parameters of the Earth's atmosphere in which the aircraft is moving. For the simulation models developed, it provides values for temperature and air density as a function of flight altitude. The task of the airframe module is to calculate the required torque that should be produced by the propulsion system. To determine it, the airframe module uses flight data from the flight control module, the aerodynamic characteristics of the aircraft and the air density depending on the altitude at which the aircraft is at any given time during the simulation from the environment module. From these, the four main forces acting on the aircraft are calculated: weight, thrust, lift and drag. The last module in the simulation model developed is propulsion, which represents both the propulsion system and energy storage. It is used to determine the relevant operating parameters of the main subsystems of the propulsion and energy storage system. This is achieved by using a backward approach that causes the propulsion module to be "forced" by the airframe module to produce the required torque, affecting the energy consumption of the electric motor and/or combustion engine.
The definition of the required operating parameters of the main subsystems of the alternative propulsion system allows for example the selection of commercially available subsystems, while also enabling the determination of the weight of the new propulsion system and the size of the main subsystems. In this way, the model can support the design process of an aircraft with alternative propulsion.
System models for propulsion provide a deeper understanding of both existing and newly designed systems. Their application in the system design process improves decision-making by enabling virtual testing and evaluation across various operational scenarios, reducing potential errors in the system and, in turn, minimising the need for physical prototypes and experiments. This results in a more efficient development process, along with reduced project costs.

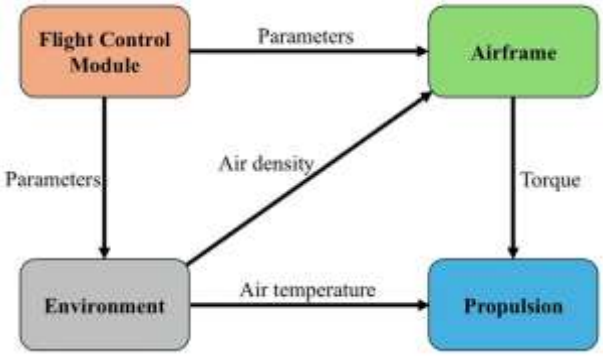


Figure 9. General diagram of an alternative propulsion system model

3 Summary

The digital environment CHESCO is developing enables the integration of external models, such as surrogate models, provided by stakeholders, thereby strengthening collaboration and jointly deepening expertise in the field of (hybrid) electric aircraft propulsion systems.
CHESCO's digital capabilities in creating an electrified aviation ecosystem are complemented by test stands being built for its main components, such as electric motors. This will allow to fine-tune simulation models and test the developed components.
Advanced research currently being conducted at CHESCO, along with growing and essential capabilities for data collection and utilisation throughout the research cycle and product life cycle, aim to create a powerful digital ecosystem that will then be used to enable choices regarding the industrialisation of electric aviation.

4 Literature

[1] Stiebe, M. Accelerating Sustainable Electric General Aviation with Fast-Charging Networks at Regional Airfields: Opportunities and Challenges. World Electr. Veh. J. 2025, 16, 118. https://doi.org/10.3390/wevj16030118.

[2] Peciak, M.; Skarka, W. Assessment of the Potential of Electric Propulsion for General Aviation Using Model-Based System Engineering (MBSE) Methodology. Aerospace 2022, 9, 74. https://doi.org/10.3390/aerospace9020074.