CONFERENCE SERIES

Johannes Vorwerk, Michael Handler, Daniel Baumgarten (Eds.)

# Proceedings of the Workshop Biosignale 2026

February 25-27, 2026, Innsbruck, Austria

*innsbruck* university press

CONFERENCE SERIES

*innsbruck* university press

universität innsbruck

Johannes Vorwerk, Michael Handler, Daniel Baumgarten (Eds.)

# Proceedings of the Workshop Biosignale 2026

February 25-27, 2026, Innsbruck, Austria

Johannes Vorwerk
Michael Handler
Daniel Baumgarten
Biomedical Engineering Group, Department of Mechatronics, University of Innsbruck, Innsbruck, Austria



© *innsbruck* university press, 2026
University of Innsbruck
1st edition
All rights reserved.
*innsbruck* university press, Karl-Schönherr-Straße 3, 6020 Innsbruck
Tel. +43 512 507-31700, iup@uibk.ac.at, www.uibk.ac.at/iup
ISBN 978-3-99106-194-6
DOI 10.15203/99106-194-6
Creative Commons Attribution 4.0 International (CC BY 4.0)

# Table of Contents

 DOI 10.15203/99106-194-6-01

# Cuff pressure control for continuous blood pressure monitoring

Sami Trimech, Christof Röhrig

***Abstract*—The precise and continuous acquisition of arterial blood pressure is essential for cardiovascular diagnostics and the long-term monitoring of at-risk patients. Conventional oscillometric cuff systems operate with short-term inflation pressures of approximately 150–200 mmHg (systolic pressure), which limits their suitability for comfortable, long-duration applications [1].**

**This paper presents a newly developed pressure-regulated cuff system that serves as a technological foundation for future AI-assisted methods of indirect blood pressure estimation. The system employs low, patient-friendly contact pressures designed to optimize the mechanical coupling between the skin and the underlying artery.**

**The hardware comprises a miniature pump, an electromagnetic valve, and a sensor unit for continuous monitoring of the cuff pressure. Prior to each measurement, a target contact pressure is defined and maintained continuously through a closed-loop control mechanism. By systematically varying the contact pressures, an optimal range was identified that provides stable, reproducible, and low-noise signals.**

**These results establish the technical basis for subsequent reference studies and for the development of data-driven, AI-based models for indirect arterial blood pressure estimation.**

## I. Introduction

The measurement of arterial blood pressure is a central component of cardiovascular diagnostics and the prevention of cardiovascular diseases, which represent the leading cause of death worldwide [1], [2]. Precise and continuous blood pressure monitoring enables the early detection of hypertension and hypotension, the optimization of therapeutic interventions, and the prevention of secondary complications such as stroke or myocardial infarction. While invasive measurement techniques provide direct and continuous blood pressure assessment, they are associated with risks such as infections, thrombosis, and increased patient invasiveness, making them unsuitable for long-term monitoring [3], [4], [5].

Conventional oscillometric cuff systems offer a non-invasive alternative, but they require short-term inflation pressures, typically ranging from 150 to 200 mmHg, corresponding to the systolic pressure [6], which limits comfort and restricts their suitability for continuous use. Recent approaches aim to extract blood pressure–related information indirectly using mechanical sensors and to process these signals with data-driven methods, particularly artificial intelligence. Piezoelectric sensor elements have the potential to capture pressure fluctuations caused by the arterial pulse directly at the wrist or upper arm [7]. A key challenge, however, lies in achieving optimal mechanical coupling between the sensor, the skin, and the artery: only under appropriate contact pressures can stable, low-noise, and reproducible signals be captured that can reliably support subsequent blood pressure estimation.

This work describes an already developed pressure-regulated cuff actuator system designed to generate low, patient-friendly contact pressures. The objective was to identify a pressure range in which piezoelectric sensors yield stable and reproducible pulse modulation signals. The developed actuator forms the technological basis for reference studies as well as for AI-assisted models that indirectly estimate arterial blood pressure from these signals. This enables continuous blood pressure monitoring without relying on high cuff pressures or invasive measurement techniques.

## II. Methods

For the present investigation, a custom-developed pressure control unit is employed that generates and continuously maintains a low and stable contact pressure to ensure precise operation of the piezoelectric sensors integrated into the cuff. The hardware comprises a proportionally driven pump, a venting valve, and an integrated pressure sensor, which together form a closed-loop control system enabling reliable pressure stabilization (Fig. 2) [8].

The cuff contains three piezoelectric sensors mounted in a custom-fabricated silicone holder (Fig. 1). This holder ensures precise sensor positioning during measurement and enables stable and reliable detection of pressure waves generated by arterial pulsation and modulation of the cuff pressure.

The use of three sensors allows spatially differentiated acquisition of the pressure waves along the cuff, reducing position-dependent variations and improving robustness and reproducibility of the measured pulse wave signals.

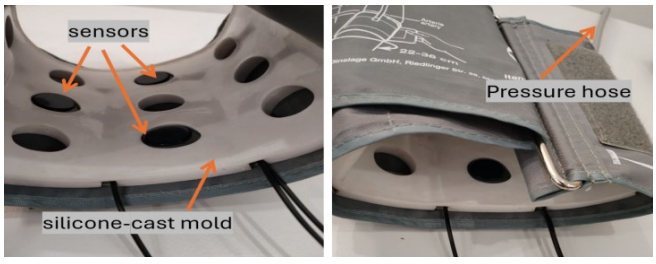


Fig. 1: Blood pressure cuff with piezoelectric sensors [8]

[1] S. Trimech, C. Röhrig are with the Department of Fachhochschule Dortmund, Institute for the Digital Transformation of Application and Living Domains (IDiAL), Sonnenstr. 96, sami.trimech@fh-dortmund.de

 DOI 10.15203/99106-194-6-01

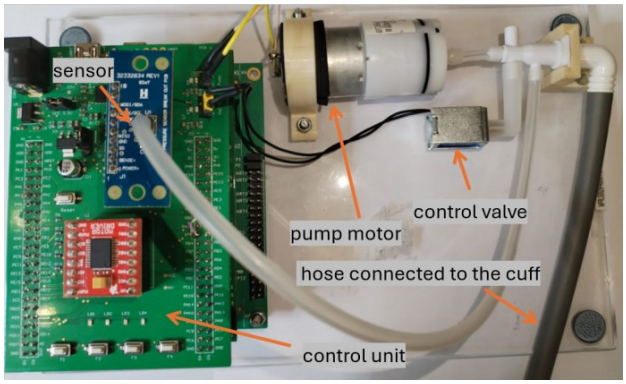


Fig. 2: Pressure regulator unit [8]

### A. Pressure Control

Before each measurement, a defined target contact pressure is set via the software interface and can be adjusted in steps within a range of 10 to 60 mmHg. The control algorithm continuously compares the measured cuff pressure with this target value and dynamically adjusts the pressure using the pump and valve.

The closed-loop control attempts to maintain the contact pressure at the set target value, regardless of external disturbances such as arm movements or changes in the participant's position. The signal quality of the recorded pressure waves thus remains largely stable throughout the measurement. Data recorded during strong arm movements that cause significant deviations will be discarded during analysis to ensure reproducible and valid results. For safety reasons, the maximum cuff pressure is limited to 200 mmHg.

### B. Signal Acquisition

Under stabilized contact-pressure conditions, the three piezoelectric sensors integrated into the cuff continuously record the pressure waves generated by arterial pulsation. The custom-made silicone holder ensures that the sensors remain precisely in their designated positions, which is essential for consistent and accurate measurements. Three datasets are collected during each measurement session:

- Data from the developed system: This includes the continuously measured actual contact pressure as well as the predefined target contact pressure. These values enable the analysis of the system's transmission behavior and the assessment of signal stability under different pressure conditions.
- Data from the piezoelectric sensors: The pressure-based pulse signals captured by the three sensors are used directly as input for the subsequent arterial blood pressure estimation.
- Data from the Finapres® NOVA system: These datasets provide complementary information on arterial pressure waveforms and serve as a reference for validating the blood pressure estimates.

All acquired data are combined to reliably derive blood pressure values based on the measured pressure waves.

### C. Internal Study Phase

At the beginning of the project, an internal study phase was conducted to test the developed system under controlled conditions and to ensure its readiness for clinical application. This phase allowed us to identify potential errors, optimize the system, and collect initial data. Building on these results, the pilot study was then carried out with 11 participants, each undergoing three measurements under the same parameters to test the system in practice and to ensure traceability and consistency in preparation for the subsequent clinical study.

During this initial phase, the contact pressure generated by the system was stabilized, and the signal acquisition of the piezoelectric sensors was evaluated. Various testing conditions and measurement parameters were examined to ensure that the system delivers reliable, reproducible, and stable measurements. The primary objectives were to validate the functionality of the pressure control mechanism and to optimize sensor data acquisition across different levels of coupling pressure.

This internal study phase also included an examination of system stability under various external influences, such as arm movements or changes in body position, to verify that measurements remain reliable under real-world conditions.

The results of this phase provided crucial information for further system optimization and established the foundation for the next step of the study, which includes comparison with established reference systems and clinical validation.

## III. RESULTS

The pressure control unit demonstrated stable and dynamically robust regulation performance in all experiments. The predefined target contact pressure was maintained nearly constant throughout the entire measurement period. Pressure deviations were compensated within a short time, even during slight arm movements.

Fig. 3 illustrates the progression of the target and actual contact pressure. The diagram shows only minimal deviations from the target value and a rapid return to the desired pressure following transient disturbances.

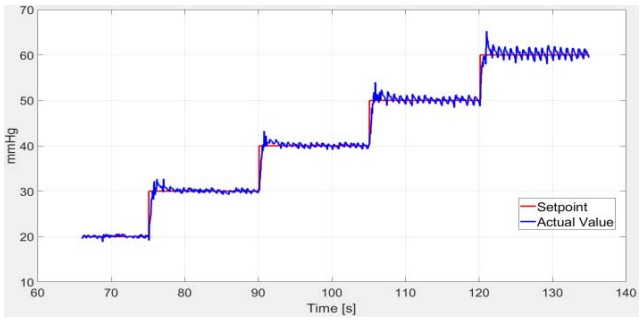


Fig. 3: Temporal profile of target and measured pressure

Under stable contact-pressure conditions, consistent and well-structured pressure-wave signals were recorded. The presented curves show clearly defined pulse peaks and a low noise level. With increasing contact pressure, the oscillation of

the pulsed components becomes more pronounced, closely correlating with the rhythmic fluctuations of the heartbeat. These oscillations reflect the heart-rate–dependent modulation of blood pressure and demonstrate the synchronization of the pressure waves with arterial pulsation.

Fig. 4 shows the measured pressure wave signals from the three piezoelectric sensors at a cuff pressure of 40 mmHg, each positioned at different locations on the cuff. The diagrams illustrate how the pulse pressure waves vary depending on the sensor position. It can be observed that signal amplitude differs across positions, indicating the local mechanical response of the artery and surrounding tissue.

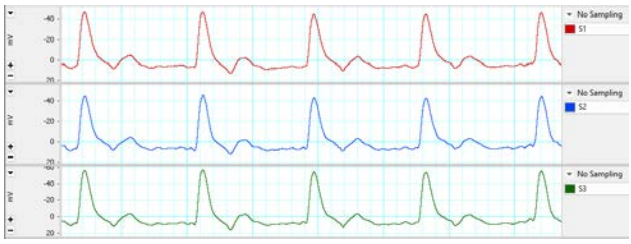


Fig. 4: Piezoelectric sensor signals

## IV. DISCUSSION

Internal tests confirm that the developed pressure control unit generates stable contact pressures, thereby producing consistent pulse signals. The results highlight the direct impact of contact pressure on signal quality and underscore the importance of precise pressure regulation for reliable acquisition of arterial pressure waves. In particular, the data show that within a defined pressure range, signals are stable, reproducible, and exhibit minimal noise characteristics that are critical for subsequent use in data-driven and AI-based blood pressure estimation models.

As the project is currently in an early developmental stage, the present findings are based on a limited number of internal measurements. Comprehensive validation will require additional data collection involving a larger participant cohort and a range of physiological conditions.

Nonetheless, the current results provide a robust foundation for the next stage of development. They enable targeted adjustment of system parameters, optimization of sensor positioning, and preparation for future formal validation and clinical studies. Moreover, these insights offer valuable guidance for the further development of non-invasive, comfortable, and continuous blood pressure monitoring systems.

## V. CONCLUSION

This study presents a newly developed pressure control unit designed to reliably capture arterial pressure waves under low and stable contact pressures. The hardware integrates a miniature pump, an electromagnetic valve, and a pressure sensor, forming a closed-loop system that continuously monitors and adjusts cuff pressure. Results from the internal test phase demonstrate that the control loop maintains the set contact pressure precisely at the target value and effectively compensates for short-term external disturbances. This ensures consistent measurement conditions, which are essential for the reliable acquisition of pulsed pressure signals.

The internal measurement series confirmed the functionality of the overall system and showed that the piezoelectric sensors embedded in the cuff generate clearly identifiable pulse waves. Collected data include the target contact pressure, the actual measured cuff pressure, and the pulsed signals from the three sensors. These datasets provide a comprehensive overview of the system's dynamic behavior and form the basis for defining operational parameters prior to further investigations with human participants.

The presented control unit represents a central technical component for the next stage of development. It enables the acquisition of datasets that will later support the development of data-driven and AI-based methods for indirect blood pressure estimation. Future work will focus on expanding the measurement data, optimizing sensor positioning and coupling, and preparing formal validation studies to further evaluate the system's potential for comfortable, continuous, and non-invasive blood pressure monitoring.

## ACKNOWLEDGMENT

This research is supported by funding from the German Federal Ministry of Research, Technology and Space BMFTR (project number 13FH564KX0).

## REFERENCES

[1] T. G. Pickering *u. a.*, „Recommendations for Blood Pressure Measurement in Humans and Experimental Animals: Part 1: Blood Pressure Measurement in Humans: A Statement for Professionals From the Subcommittee of Professional and Public Education of the American Heart Association Council on High Blood Pressure Research“, *Hypertension*, Bd. 45, Nr. 1, S. 142–161, Jan. 2005, doi: 10.1161/01.HYP.0000150859.47929.8e.

[2] D. Panagiotakos, C. Antza, und V. Kotsis, „Ambulatory and home blood pressure monitoring for cardiovascular disease risk evaluation: a systematic review and meta-analysis of prospective cohort studies“, *J Hypertens*, Bd. 42, Nr. 1, S. 1–9, Jan. 2024, doi: 10.1097/HJH.0000000000003557.

[3] R. Shan *u. a.*, „Early blood pressure assessment after acute myocardial infarction: Insights using digital health technology“, *Am J Prev Cardiol*, Bd. 3, S. 100089, Sep. 2020, doi: 10.1016/j.ajpc.2020.100089.

[4] P. M. Mariano-Gomes, A. Ouverney-Braz, und G. Oroski-Paes, „Adverse events with arterial catheters in intensive care units: a scoping review“, *Enferm Intensiva (Engl Ed)*, Bd. 35, Nr. 4, S. 410–427, 2024, doi: 10.1016/j.enfie.2024.06.001.

[5] M. Wijnberge, B. Van Der Ster, A. P. J. Vlaar, M. W. Hollmann, B. F. Geerts, und D. P. Veelo, „The Effect of Intermittent versus Continuous Non-Invasive Blood Pressure Monitoring on the Detection of Intraoperative Hypotension, a Sub-Study“, *JCM*, Bd. 11, Nr. 14, S. 4083, Juli 2022, doi: 10.3390/jcm11144083.

[6] C. F. Babbs, „Oscillometric measurement of systolic and diastolic blood pressures validated in a physiologic mathematical model“, *Biomed Eng Online*, Bd. 11, S. 56, Aug. 2012, doi: 10.1186/1475-925X-11-56.

[7] S. Min *u. a.*, „Clinical Validation of a Wearable Piezoelectric Blood-Pressure Sensor for Continuous Health Monitoring“, *Adv Mater*, Bd. 35, Nr. 26, S. e2301627, Juni 2023, doi: 10.1002/adma.202301627.

[8] S. Trimech, C. Wolff, and C. Röhrig, "Non-invasive blood pressure measurement with piezoelectric sensors," in Proceedings of the 2025 IEEE 13th International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS), Gliwice, Germany, Sep. 2025.,

 DOI 10.15203/99106-194-6-02

# Forward Modelling in Magnetospinography

Maike Schmidt, George O'Neill, Meaghan E Spedden, Gareth Barnes

*Abstract*—Magnetospinography (MSG) enables non-invasive spinal electrophysiology through the measurement of magnetic fields generated by neural activity. As with other bio magnetic modalities, such as magnetoencephalography, accurate forward modelling is essential for reliable interpretation. In this study we perform a comparative analysis of two forward modelling techniques: Finite element modelling (FEM) and boundary element modelling (BEM) using a series of increasingly complex anatomical models. Under homogeneous conditions, FEM and BEM yield highly consistent leadfields (correlation coefficients > 0.95), indicating robustness across numerical formulations. However, deviations emerge in more anatomically intricate regions, and in caudal spinal regions, where the curvature and heterogeneity of the vertebral column becomes more pronounced. Comparisons between anatomically registered and canonically aligned geometries further reveal that uncertainty in the location of the spinal cord results in larger error values, even when comparing the same bone type between anatomically registered and canonically registered models, than when the spinal cord location is anatomically precise and different bone models are being compared. These results highlight the critical role of anatomical fidelity and alignment in MSG forward modelling and motivate standardized, anatomically grounded pipelines for emerging spinal and brain–spine biomagnetic imaging.

***Index Terms*— Magnetospinography, Forward modelling, FEM, BEM, OPM, Spinal cord, Source localization**

## I. Introduction

Magnetospinography (MSG) provides a unique avenue for non-invasive spinal electrophysiology using optically pumped magnetometers (OPMs) [1]. Accurate forward modelling is critical for source reconstruction, yet MSG's elongated geometry, heterogeneous bone structure, and variable CSF thickness pose unique challenges [2,3]. While prior work [4] demonstrated feasibility with simplified models, the effects of anatomical complexity and registration methods remain poorly characterized. This study systematically evaluates finite element modelling (FEM) and boundary element modelling (BEM) across multiple bone models and registration approaches.

*Important notes:*
The focus of this study is not just numerical accuracy but practical viability: how model choices influence localization error, cross-site compatibility, and suitability for concurrent brain-spine MSEG studies.

## II. Methods

We constructed a series of nested anatomical models derived from subject-specific MRI, including soft tissue, CSF, spinal cord and multiple bone representations:

- 
  Continuous bone model (smooth, non-segmented vertebral wrap)
- 
  Homogenous toroidal model (toroidal segments with uniform thickness and segment height)
- 
  Inhomogeneous toroidal model (toroidal segments with increasing thickness and segment height with each descending segment)
- 
  MRI derived model (model segmented from subject MRI)

Models were registered via two methods: *anatomical* (MRI-derived with torso, spinal cord and bone directly extracted) and *canonical* (fiducial-aligned to a standard torso model).

Anatomical models were derived from whole-spine MRI. Vertebral bone and spinal cord structures were segmented using TotalSegmentator [5]. Soft tissue compartments, heart and lungs from ECGSim [6], were registered into the model through Procrustes alignment followed by iterative closest point (ICP). Spinal cord dipoles were placed at regular 5mm intervals along the cord with three orthogonal orientations.

Forward solutions used BEM (Helsinki BEM Framework [7,8] with linear collocation) modelling tissue as piecewise-homogeneous regions bounded by nested surfaces (torso, lungs, heart, spinal cord, vertebral bone), and FEM (DUNEuro [9], continuous Galerkin formulation with tetrahedral meshes, St. Venant approach for dipole sources). Identical dipole locations,

M. Schmidt and all co-authors are with the Department of Imaging Neuroscience, UCL, United Kingdom, e-mail: skgtchm@ucl.ac.uk.

 DOI 10.15203/99106-194-6-02

orientations, and sensor configurations were used across methods. Leadfields were evaluated via Relative Error (RE) and Correlation Coefficient (CC):

$$RE = \frac{|\boldsymbol{L_1} - \boldsymbol{L_2}|}{|\boldsymbol{L_1}|}$$

$$CC = \left(\frac{\sum_{i=1}^{N}(L_{1,i} - \overline{L_1})(L_{2,i} - \overline{L_2})}{\sqrt{\sum_{i=1}^{N}\left(L_{1,i} - \overline{L_1}\right)^2 \sum_{i=1}^{N}\left(L_{2,i} - \overline{L_2}\right)^2}}\right.$$

We note that (for zero-mean data) alternative error measures could be used here such as the relative difference measure (RDM) or magnitude error. We report values of RE and CC as percentages, for this the results of equations 1 and 2 were multiplied by 100.

## III. Results

We evaluated two primary sources of forward modelling variability: (i) differences between numerical frameworks (FEM versus BEM) under matched anatomical conditions, and (ii) differences arising from anatomical registration (canonical versus participant-specific) within a fixed framework.

For the FEM-BEM comparison, identical bone models were embedded within the anatomically registered torso, isolating framework-dependent effects. For continuous bone, both methods showed near-perfect agreement (RE=2.5%, CC=99.8%), indicating convergence under simplified geometric assumptions. Introducing segmented vertebral geometry led to substantial divergence: the inhomogeneous toroidal model showed RE=41% and CC=72.9%. The largest discrepancies occurred with the MRI-derived realistic bone model (RE=44.4%, CC=58.2%).

Within-framework comparisons revealed markedly different stability. BEM exhibited high consistency across all bone models, with all RE values below 19% and CC values above 96%. In contrast, FEM showed greater variability (RE=11-23%, CC=88-98%), indicating higher sensitivity to increasing bone geometry complexity.

For registration effects, we compared canonical and anatomical torso models using BEM only, ensuring observed differences reflected spinal cord position rather than numerical instability. Agreement was poor across all bone representations: continuous bone showed RE=61.9% with CC=19.4%, while segmented models showed modest improvement but remained inconsistent (homogeneous: RE=50%, CC=36.6%; inhomogeneous: RE=48.2%, CC=39%).

Figures 1-4 summarize these comparisons.

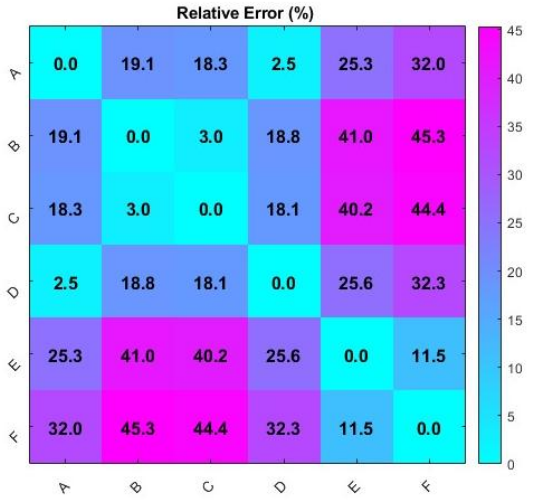


Figure 1) RE of leadfields, comparing the different bone models across the BEM and the FEM forward modelling frameworks. A: BEM, Continuous bone, B: BEM Toroidal Bone, C: BEM, Realistic Bone, D: FEM, Continuous Bone, E: FEM, Toroidal Bone, F: FEM, Realistic Bone

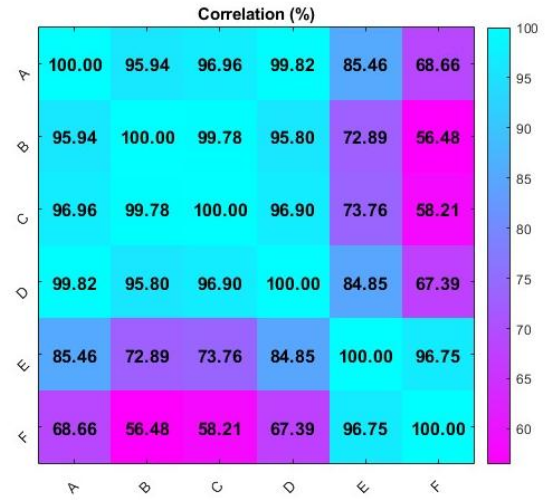


Figure 2) CC of leadfields, comparing the different bone models across the BEM and the FEM forward modelling frameworks. A: BEM, Continuous bone, B: BEM Toroidal Bone, C: BEM, Realistic Bone, D: FEM, Continuous Bone, E: FEM, Toroidal Bone, F: FEM, Realistic Bone

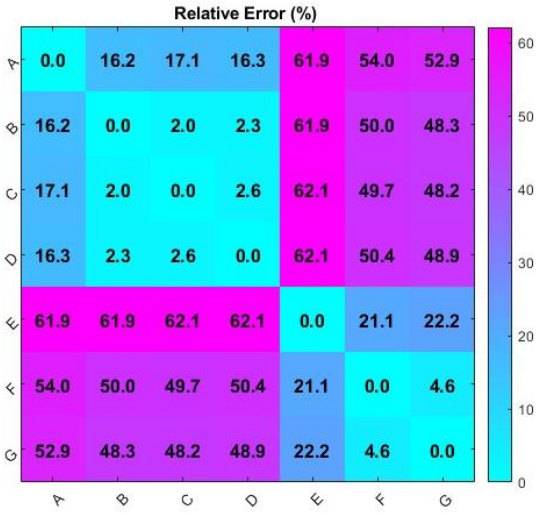


Figure 3) RE of leadfields produced using the BEM framework, comparing the different bone models across anatomical (An) and canonical (Ca) registration methods. A: Anatomical, Continuous, B: Anatomical, Homogenous, C: Anatomical, Inhomogeneous, D: Anatomical, Realistic, E: Canonical, Continuous, F: Canonical, Homogenous, G: Canonical, Inhomogeneous

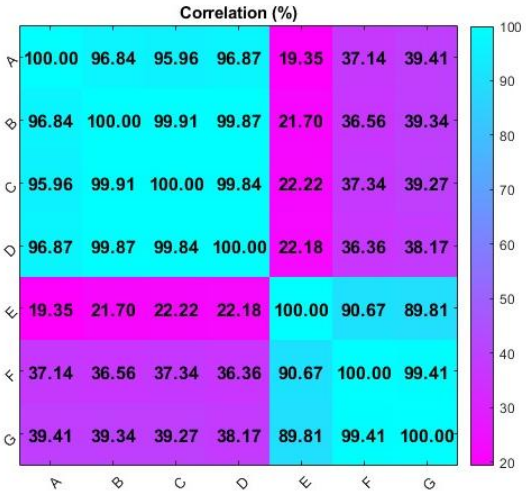


Figure 4) CC of leadfields produced using the BEM framework, comparing the different bone models across anatomical (An) and canonical (Ca) registration methods A: Anatomical, Continuous, B: Anatomical, Homogenous, C: Anatomical, Inhomogeneous, D: Anatomical, Realistic, E: Canonical, Continuous, F: Canonical, Homogenous, G: Canonical, Inhomogeneous

## IV. Discussion and Limitations

These results demonstrate that forward modelling accuracy in MSG is shaped by two interacting factors: numerical formulation and anatomical registration.

FEM and BEM converge under simplified geometry but diverge with realistic vertebral structure. BEM smooths local variations via surface integrals, while FEM resolves volumetric current flow, increasing sensitivity to anatomical detail. Critically, anatomical registration errors exceed bone model effects: even with stable numerical evaluation, cord position errors produce large leadfield changes. This suggests participant-specific anatomy may be prerequisite for meaningful MSG interpretation, though validation with experimental data remains needed.

These results suggest a practical hierarchy for MSG modelling. Simplified geometries and canonical anatomy may be sufficient for exploratory or group-level analyses, where individual anatomical variability is averaged out. However, participant-specific anatomy may be required for studies aiming to interpret source location, orientation, or segment-specific activity. FEM may offer advantages when fine anatomical detail is available and explicitly required, whereas BEM provides greater robustness when such detail is uncertain or unavailable.

Several limitations should be noted. isotropic conductivity assumptions (known anisotropy in white matter would increase FEM sensitivity) and focus on forward metrics rather than inverse localization performance

## V. Conclusion

Both FEM and BEM are viable under simplified conditions, but anatomically realistic models are essential for complex spinal regions. Anatomical registration accuracy dominates forward model performance, with implications for standardized protocols in concurrent brain-spine MSEG.

## References

[1] L. C. Lydia, “Concurrent spinal and brain imaging with optically pumped magnetometers,” *Journal of Neuroscience Methods*, vol. 406, p. 110131, 2024, doi: 10.1016/j.jneumeth.2024.110131.

[2] D. W. Cadotte, A. Cadotte, J. Cohen-Adad, D. Fleet, M. Livne, J. R. Wilson, D. Mikulis, N. Nugaeva, and M. G. Fehlings, "Characterizing the location of spinal and vertebral levels in the human cervical spinal cord," *American Journal of Neuroradiology*, vol. 36, no. 4, pp. 803–810, 2015.

[3] M. Kaptan, D. Pfyffer, C. G. Konstantopoulos, C. S. W. Law, K. A. Weber II, G. H. Glover, and S. Mackey, “Recent developments and future avenues for human corticospinal neuroimaging,” *Frontiers in Human Neuroscience*, vol. 18, p. 1339881, Jan. 2024.

[4] O’Neill, G.C., Spedden, M.E., Schmidt, M. *et al.* Volume conductor models for magnetospinography. *Sci Rep* **15**, 26258 (2025). https://doi.org/10.1038/s41598-025-10770-z

[5] Yehuda Warszawer, Nathan Molinier, Jan Valosek, Pierre-Louis Benveniste, Sandrine B´edard, Emanuel Shirbint, Feroze Mohamed, Charidimos Tsagkas, Shannon Kolind, Larry Lynd, Jiwon Oh, Alexandre Prat, Roger Tam, Anthony Traboulsee, Scott Patten, Lisa Eunyoung Lee, Anat Ach-iron, and Julien Cohen-Adad. Totalspineseg: Robust spine segmentation with landmark-based labeling in mri. ResearchGate preprint, 2025. URL https://www.researchgate.net/publication/389881289_TotalSpineSeg_Robust_Spine_Segmentation_with_Landmark-Based_Labeling_in_MRI

[6] van Oosterom A, Oostendorp TF. ECGSIM: an interactive tool for studying the genesis of QRST waveforms. Heart 90, 165-168 (2004).

[7] M. Stenroos and J. Sarvas. Bioelectromagnetic forward problem: Isolated source approach revis(it)ed. Physics in Medicine and Biology, 57(11), 2012. ISSN 00319155. doi: 10.1088/0031-9155/57/11/3517

[8] J. Sarvas. Basic mathematical and electromagnetic concepts of the biomagnetic inverse problem. Physics in Medicine and Biology, 32(1), 1987. ISSN 00319155. doi:10.1088/0031-9155/32/1/004.

[9] S. Schrader, A. Westhoff, M. C. Piastra, T. Miinalainen, S. Pursiainen, J. Vorwerk, H. Brinck, C. H. Wolters, and C. Engwer. DUNEuro – a software toolbox for forward modeling in bioelectromagnetism. PLOS ONE, 16(6):e0252431, 2021. doi: 10.1371/journal.pone.0252431. URL https://doi.org/10.1371/journal.pone.0252431.

 DOI 10.15203/99106-194-6-03

# Quantification of Magnetic Nanoparticles Using a Surface Acoustic Wave Magnetoelastic Sensor

L. M. W. Bangert, G. Barbieri, F. Weisheit, H. Wolframm, M. A. Nowak, D. Meyners, J. McCord, M. Höft, E. Quandt, M. Gerken

***Abstract*—Magnetic nanoparticles (MNPs) play a key role in biomedical technologies, including targeted drug delivery, contrast imaging, and protein detection. In this work, the detection capability of a surface acoustic wave (SAW) magnetoelastic (ME) sensor for $Fe_2O_3$ MNPs was investigated under unshielded conditions. The study focuses on detecting 50 µL of different concentrations of MNP suspensions ranging from 10 $\mathrm{mg\,mL^{-1}}$ to 25 $\mathrm{mg\,mL^{-1}}$. The magnetic excitation was achieved using an alternating magnetic field of 100 µT at a frequency of 923.67 Hz, and the resulting magnetization changes were transduced into phase shifts of the SAW. The observed decrease in the out-of-phase component of the magnetic susceptibility with decreasing MNP concentration is consistent with the theoretical expectation obtained by the Debye model. The AC susceptibility measurements using the SAW ME sensor demonstrate reliable MNP detection in an unshielded environment, providing an important step toward compact, contact-free magnetic biosensing platforms.**

***Index Terms*—AC susceptibility, biosensing, Brownian relaxation, magnetic nanoparticles, surface acoustic wave sensor**

## I. Introduction

Magnetic nanoparticles (MNPs) are widely used in targeted drug delivery, magnetic resonance imaging, and biosensing via functionalized magnetic labels [1]. Accurate detection and quantification of MNPs is crucial for controlling dosage, localization, and binding efficiency, particularly in in-vivo diagnostic systems where real-time monitoring is required to assess particle transport and clearance [2]. Compact, sensitive and unshielded sensing approaches are therefore necessary to support clinical translation.

Existing magnetic sensing methods include Hall effect and magnetoresistive sensors [3], [4], which offer miniaturization but are limited by 1/f noise and thermal drift in low-field environments [5]. Fluxgate and optically pumped magnetometers allow nanotesla-range measurements suitable for laboratory characterization of MNPs [6], [7].

Surface acoustic wave (SAW) devices present an attractive alternative due to their high bandwidth, compact size, and compatibility with lab-on-chip integration [8], [9]. When combined with a magnetoelastic (ME) layer, SAW devices exhibit a phase shift under magnetic excitation [10]. As SAW ME sensors are highly sensitive to small magnetic field changes, they are promising for detecting MNP and their concentrations. However, to the best of the authors' knowledge, direct, unshielded MNP detection using SAW ME sensors has not yet been demonstrated. In this work, we evaluate their feasibility for detecting different concentrations of encapsulated $Fe_2O_3$ MNPs.

This work was supported by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG) through Collaborative Research Center CRC 1261 (Project-ID 286471992). Corresponding authors: L. M. W. Bangert, M. Gerken (e-mail: {luba, mge}@tf.uni-kiel.de).

All authors are with the Kiel University, Kiel, Germany.

## II. Methods

### A. Rotational Dynamics of Magnetic Nanoparticles

The rotational dynamics of MNPs are generally governed by two mechanisms: Brownian relaxation, which arises from the physical rotation of the whole particle in a fluid, and Néel relaxation, corresponding to the internal flipping of the particle's magnetic moment [11]. For thermally blocked particles, the effective relaxation is dominated by Brownian motion, whose characteristic time $\tau_\mathrm{B}$ is given by

$$\tau_\mathrm{B} = \frac{3\eta V_\mathrm{H}}{k_\mathrm{B}T}, \tag{1}$$

where $\eta$ is the dynamic viscosity of the medium, $V_\mathrm{H}$ the hydrodynamic volume of the particle, $k_\mathrm{B}$ the Boltzmann constant, and $T$ the absolute temperature [6].

The magnetic AC susceptibility (ACS) characterizes the response of an ensemble of MNPs to an oscillating magnetic field. When the excitation frequency $f$ is comparable to the magnetic relaxation rate, the system exhibits a phase lag between the applied field and the resulting magnetization, leading to energy dissipation. Based on the Debye model, the dynamic magnetic susceptibility is therefore expressed as a complex quantity

$$\chi(\omega) = \chi'(\omega) - i\,\chi''(\omega), \tag{2}$$

where $\omega = 2\pi f$ denotes the angular excitation frequency. The real part $\chi'(\omega)$ describes the in-phase, reversible magnetic response, while the imaginary part $\chi''(\omega)$ captures the out-of-phase component associated with dissipative losses [12].

For Brownian-dominated relaxation, it follows:

$$\chi' = \frac{\chi_0}{1+(\omega\tau_\mathrm{B})^2}, \qquad \chi'' = \frac{\chi_0\,\omega\tau_\mathrm{B}}{1+(\omega\tau_\mathrm{B})^2}. \tag{3}$$

The static susceptibility $\chi_0$ of the particle ensemble is expressed as

$$\chi_0 = \frac{\mu_0 N \mu^2}{3k_\mathrm{B}T}, \tag{4}$$

with $\mu_0$ the vacuum permeability, $N$ the particle number density and $\mu$ the magnetic moment of a single particle. The maximum of $\chi''(\omega)$ occurs at $\omega = 1/\tau_\mathrm{B}$, corresponding to the frequency of maximum energy dissipation. Measuring $\chi'(\omega)$ and $\chi''(\omega)$ thus provides direct insight into the hydrodynamic

 DOI 10.15203/99106-194-6-03

size, viscosity, magnetic moment, and concentration of the MNPs [6], [11].

### B. AC Susceptibility Measurement Setup

The measurements were done using the same unshielded and portable setup described by Barbieri *et al.* [13]. An ACS technique was employed to excite the thermally blocked $Fe_2O_3$ MNP using an AC magnetic field of $B = 100$ µT, generated by a 25 cm diameter excitation coil shown in Fig. 1 (a) driven by a 4-quadrant voltage and current amplifier (A1110-40 QE-100, Dr. Hubert GmbH). The field strength was monitored by a Teslameter (FM302, Project Elektronik Mess- und Regelungstechnik GmbH) and the coil current $I_{\mathrm{rms}} \approx 780\,\mathrm{mA}$ adjusted accordingly. Signal generation and read-out were performed simultaneously with a 32-bit sound card (ADI2-Pro FS-R, RME) using synchronized channels to extract their phase difference, and processed in MATLAB® R2024b.

An excitation frequency of 923.67 Hz was selected as the closest available value in the calibrated setup to the frequency corresponding to $\tau_{\mathrm{B}}$ of the MNPs employed in this study. Using (1) with the dynamic viscosity of water ($\eta = 0.89$ mPa s at $T = 298$ K [14]) and the particle volume $V_{\mathrm{H}} = \frac{4}{3}\pi(d_{\mathrm{H}}/2)^3$ for the hydrodynamic diameter $d_{\mathrm{H}} = 80$ nm, it follows that $\tau_{\mathrm{B}} \approx 174$ µs, yielding a resonance frequency of $f \approx 914.8$ Hz.

In the original setup, a differential gradiometer consisting of two pick-up coils was used for magnetic field detection. In this work, the gradiometer was replaced by a SAW ME sensor positioned at the center of the excitation coil. The sensor was fabricated within the collaborative research center (CRC 1261) [15] and operated in an open-loop configuration by maintaining a dedicated operational point via a proportional-integral-derivative control loop. The sensor is excited, controlled and evaluated with a lock-in amplifier (GHFLI, Zurich Instruments). We obtain the magnetic information by analyzing the control error. The sensor is excited at a frequency of 148.6 MHz and the resulting signal is sampled at 119 kHz.

### C. Samples and Measurement Procedure

Commercial $Fe_2O_3$ MNP (Bionized NanoFerrite (BNF)-Starch, micromod Partikeltechnologie GmbH, 80 nm) with a stock concentration of 25 mg mL$^{-1}$ (14 mg mL$^{-1}$ iron) were diluted to 10–25 mg mL$^{-1}$, as shown in Fig. 1 (c). The in-phase $X$ and out-of-phase $Y$ components of the system output, corresponding to $\chi'$ and $\chi''$ of the MNP sample, were monitored over time as depicted in Figure 2. Each measurement was repeated three times and had a duration of 60 s with the sample placed on the sample holder next to the SAW ME sensor after 10 s and removed again after 40 s of measurement time $t$. For quantitative evaluation, the mean signal between $t = 20$ s and $t = 30$ s was used.

At first, 50 µL of the MNP suspensions were encapsulated in a cylindrical polycarbonate capsule with size 4 mm × 4 mm × 10 mm and placed at a distance of 1 mm from the SAW ME sensor as shown in Fig. 1 (b). Next, an empty capsule was measured to estimate the noise floor. This background signal was later subtracted from the sample measurements. Last, an encapsulated $Dy_2O_3$ reference sample (67.2 mg, $m_{\mathrm{cal}} = 1.76 \times 10^{-8}$ A m$^2$) was measured and used for calibration and normalization as described in [13].

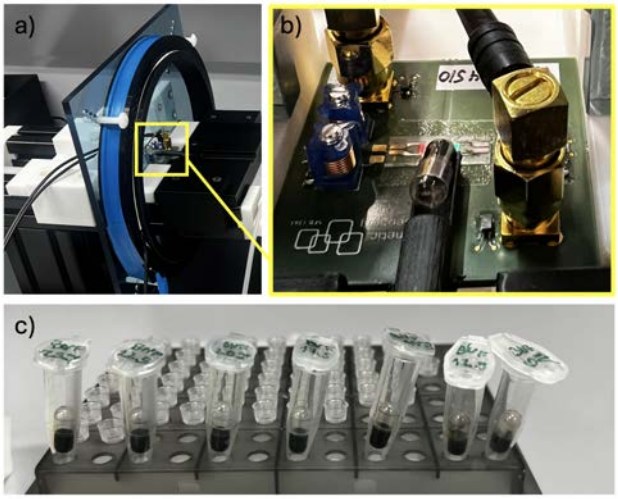


Fig. 1: a) Experimental setup, including the excitation coil and the SAW ME sensor. b) MNP sample placed next to the SAW ME sensor. c) Encapsulated MNP samples arranged from highest (left) to lowest (right) concentration.

## III. Results and Discussion

In the top right panel of Fig. 2, the sensor output $Y$ remains near 0 V until the MNP sample is introduced at $t = 10$ s, after which it rises to $Y \approx 4.7 \times 10^{-5}$ V. The signal remains at this level, with only minor fluctuations due to measurement noise, until the sample is removed at $t = 40$ s, reflecting energy dissipation associated with the Brownian relaxation of the MNPs. Hence, the observed increase confirms that the MNPs exhibit ferrimagnetic behavior.

In contrast, the $Dy_2O_3$ reference sample, depicted in the bottom right of Fig. 2 produces no measurable signal change and keeps $Y$ at the noise floor, consistent with its purely paramagnetic response under the applied alternating magnetic field. Thus, an increase in the $Y$ output reliably reflects the presence of a dynamically relaxing magnetic sample and forms the basis for extracting $\chi$.

Building on this, Fig. 3 shows the out-of-phase susceptibility $\chi''_{\mathrm{norm}}$ normalized to the reference sample as a function of MNP concentration at $f = 923.67$ Hz. As predicted by the Debye model in Eqs. (2)–(4), $\chi''_{\mathrm{norm}}$ increases with particle concentration, since $\chi_0$ scales with $N$. Across the investigated range (10–25 mg/mL), the data follow an approximately linear trend with a fitted slope of $\sim 0.018$ (mg/mL)$^{-1}$, rising from 0.18 to 0.44, and the fit is consistent with a line passing through the origin. This behavior confirms that the magneto-mechanical coupling between the ensemble of thermally blocked MNPs and the AC field increases proportionally with the MNP concentration, as expected from the theoretical model.

Minor deviations from linearity may arise from partial aggregation of MNPs or pipetting inaccuracy. Nevertheless,

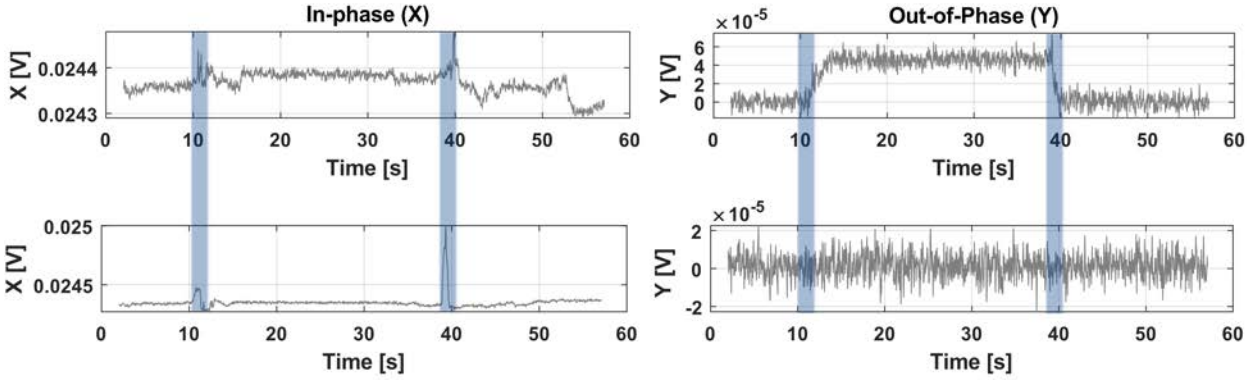


Fig. 2: Time-resolved in-phase (left) and out-of-phase (right) voltage output of the SAW ME sensor system for a $Fe_2O_3$ MNP sample (25 mg mL$^{-1}$, top) and a $Dy_2O_3$ reference (bottom), placed next to the sensor at 10 s and removed at 40 s.

the overall trend supports the theoretical prediction that $\chi''$ serves as a direct indicator of $N$ and thus MNP concentration.

## IV. Conclusion

In this study, we demonstrated that SAW ME sensors are a feasible platform for detecting different concentrations of thermally blocked $Fe_2O_3$ MNPs. The measured $\chi''$ increased with particle concentration, in agreement with theoretical predictions based on Brownian relaxation and the Debye model. These results confirm that SAW ME sensors reliably transduce the rotational dynamics of MNP ensembles into electrical signals, highlighting their potential for sensitive and compact biosensing applications in unshielded environments.

## Acknowledgment

This research was funded by the German Research Foundation (Deutsche Forschungsgemeinschaft) through project A1, A8, A9, B1, B12, Z1 and Z2 of the CRC 1261 "Magnetoelectric Sensors: From Composite Materials to Biomagnetic Diagnostics".

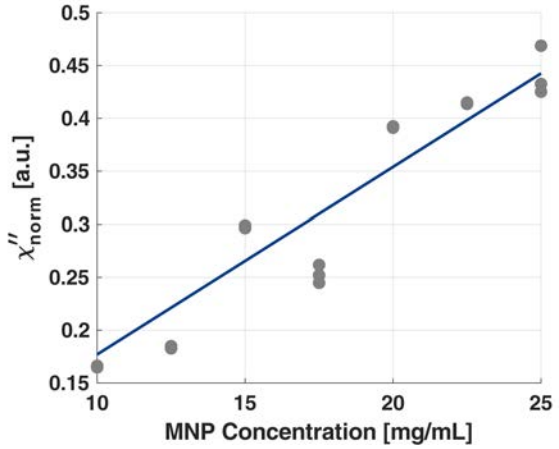


Fig. 3: Normalized out-of-phase susceptibility versus MNP concentration at $f = 923.67$ Hz and $B = 100$ µT.

## References

[1] K. M. Krishnan, "Biomedical nanomagnetics: A spin through possibilities in imaging, diagnostics, and therapy," *IEEE Transactions on Magnetics*, vol. 46, no. 7, pp. 2523–2558, 7 2010.

[2] M. Liebl, F. Wiekhorst, D. Eberbeck, P. Radon, D. Gutkelch, D. Baumgarten, U. Steinhoff, and L. Trahms, "Magnetorelaxometry procedures for quantitative imaging and characterization of magnetic nanoparticles in biomedical applications," *Biomedizinische Technik*, vol. 60, no. 5, pp. 427–443, 10 2015.

[3] G. Mihajlović, P. Xiong, S. Von Molnár, M. Field, G. J. Sullivan, K. Ohtani, and H. Ohno, "Submicrometer hall sensors for superparamagnetic nanoparticle detection," *IEEE Transactions on Magnetics*, vol. 43, no. 6, pp. 2400–2402, 2007.

[4] H. Ahn, A. Tanaka, Y. Kono, S. B. Trisnanto, T. Kasajima, T. Shibuya, and Y. Takemura, "Detection of Sub-pT Field of Magnetic Responses in Metals and Magnetic Materials by Highly Sensitive Magnetoresistive Sensors," *Sensors*, vol. 25, no. 3, 2 2025.

[5] N. J. Darton, A. Ionescu, and J. Llandro, Eds., *Magnetic Nanoparticles in Biosensing and Medicine*. Cambridge University Press, 2 2019.

[6] F. Ludwig and H. Remmer, "Rotational dynamics of magnetic nanoparticles in different matrix systems," *Physical Sciences Reviews*, vol. 7, no. 9, pp. 981–1008, 9 2022.

[7] O. Baffa, R. H. Matsuda, S. Arsalani, A. Prospero, J. R. Miranda, and R. T. Wakai, "Development of an optical pumped gradiometric system to detect magnetic relaxation of magnetic nanoparticles," *Journal of Magnetism and Magnetic Materials*, vol. 475, pp. 533–538, 4 2019.

[8] K. Länge, B. E. Rapp, and M. Rapp, "Surface acoustic wave biosensors: A review," *Analytical and Bioanalytical Chemistry*, vol. 391, no. 5, pp. 1509–1519, 7 2008.

[9] D. B. Go, M. Z. Atashbar, Z. Ramshani, and H. C. Chang, "Surface acoustic wave devices for chemical sensing and microfluidics: A review and perspective," pp. 4112–4134, 7 2017.

[10] H. Wolframm, F. Weisheit, E. Quandt, and M. Höft, "Dedicated SAW Oscillator for Sensing Applications," in *2024 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*, Glasgow, United Kingdom, 2024, pp. 1–5.

[11] J. Dieckhoff, D. Eberbeck, M. Schilling, and F. Ludwig, "Magnetic-field dependence of Brownian and Néel relaxation times," *Journal of Applied Physics*, vol. 119, no. 4, 1 2016.

[12] C. V. Topping and S. J. Blundell, "A.C. susceptibility as a probe of low-frequency magnetic dynamics," *Journal of Physics Condensed Matter*, vol. 31, no. 1, 1 2019.

[13] G. Barbieri, J. Arbustini, E. Elzenheimer, F. Wiekhorst, P. Radon, A. Teplyuk, R. Rieger, A. Bahr, M. Höft, and M. Gerken, "Unshielded Portable Gradiometer Setup for Magnetic Nanoparticles Characterization in Flow Condition," *TechRxiv*, 2025.

[14] W. M. Haynes, D. R. Lide, and T. J. Bruno, "CRC Handbook of Chemistry and Physics 97 th Edition," Tech. Rep., 2017.

[15] A. Kittmann, P. Durdaut, S. Zabel, J. Reermann, J. Schmalz, B. Spetzler, D. Meyners, N. X. Sun, J. McCord, M. Gerken, G. Schmidt, M. Höft, R. Knöchel, F. Faupel, and E. Quandt, "Wide Band Low Noise Love Wave Magnetic Field Sensor System," *Scientific Reports*, vol. 8, no. 1, 12 2018.

 DOI 10.15203/99106-194-6-04

# Characterising Redundancy Between Long Term Subcutaneous Electroencephalography and Chest Accelerometry in a Healthy Sleep Cohort

J. Verne Henriksen, J. L. Vishart, N. Spicher, J. Duun-Henriksen

***Abstract*—Subcutaneous electroencephalography (sqEEG) enables long term neural monitoring in natural home settings and has shown strong potential for sleep analysis. Some systems also record chest mounted accelerometry, but it is unclear how much complementary information this modality provides beyond the EEG signal. This study characterises redundancy between sqEEG and chest accelerometry in a healthy sleep cohort. Recording from N = 20 participants in the Ultra Long Term Sleep dataset were analysed, comprising continuous sqEEG and triaxial chest accelerometry with automated sleep stage labels. Amplitude and spectral features were extracted from sqEEG, and posture and movement features from accelerometry, in 30 second epochs. Principal Component Analysis (PCA) was applied to the combined multimodal feature set and to modality specific subsets. In the joint PCA, the first two components explained 31.6% of the variance and were not dominated by any particular modality. Accelerometry alone explained 38.1% of variance, and sqEEG alone 58.0%. The combined feature space closely resembled the sqEEG only space, indicating substantial redundancy between modalities. These findings suggest that, in healthy subjects, chest accelerometry adds limited new information beyond long term sqEEG for variance based characterisation of sleep. The framework provides a baseline for future multimodal studies in clinical populations.**

***Index Terms*—accelerometry, multimodal biosignals, principal component analysis, sleep staging, subcutaneous EEG**

## I. Introduction

Longitudinal neural monitoring is becoming increasingly relevant for studying neurological conditions such as epilepsy. Subcutaneous electroencephalography (sqEEG) systems provide stable long duration recordings for home use and retain meaningful spectral and temporal structures for neurological evaluation. This makes the modality promising for characterization of both healthy and pathological EEG activity [1].

Many inpatient EEG systems include accelerometry to capture posture or movement, for example for sleep staging. Although accelerometry has been shown to be of value for this task, its contribution when combined with sqEEG for long term studies is less clear. In particular, how much unique information a chest mounted accelerometer provides relative to the sqEEG signal [2], [3].

This work evaluates whether chest accelerometry offers complementary information beyond what can already be extracted from long term sqEEG in at-home subject settings.

J. Verne Henriksen is a student at DTU.

J. Verne Henriksen, J. L. Vishart, and J. Duun-Henriksen are with UNEEG medical A/S, Borupvang 2, 3450 Lillerød.

N. Spicher is with the Department of Healthcare Technology, Technical University of Denmark. Anker Engelunds Vej 1, 2800 Kongens Lyngby.

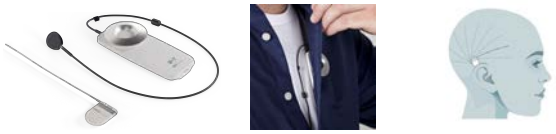

Fig. 1: The measurement setup analysed in this work. **Left:** UNEEG medical's SubQ implant and 24/7 EEG SubQ Recorder. **Center:** Demonstration of recorder placement in the upper chest area. **Right:** Subcutaneous implant and lead positioning. Images rights are with UNEEG medical A/S [5].

It uses data from an ultra long duration healthy sleep cohort and provides a methodological foundation for the multimodal analysis planned for future work [4].

## II. Material and Methods

UNEEG medical's sqEEG system (Fig. 1) uses a small two channel implant positioned behind the ear with a lead placed towards the temporal region. The implant connects to a chest worn recorder that logs and stores continuous long term EEG measurements together with triaxial accelerometry, making the system easy to use and wear during daily life.

Recordings were obtained from the Ultra Long Term Sleep dataset [4], which includes continuous sqEEG and accelerometry acquired in the subjects' home environment. Demographic information is available in Table I. Subjects attached the cable between the implant and recorder using an adhesive patch, and secured the recorder itself to their clothing during sleep.

SqEEG was sampled at 207 Hz from a two channel implant. Accelerometry was sampled at 12.5 Hz using the sensor embedded in the chest recorder. Sleep stage labels in Wake, N1 to N3, and REM were provided by an automated scoring algorithm validated for sqEEG [4], [6]. Preprocessing included artefact removal of non-physiological spikes and device related logging artefacts with linear interpolation for both sqEEG and accelerometry. sqEEG signal was already filtered with a $[0.5-48]$ Hz FIR equiripple filter with 40 dB attenuation in the stopbands. Accelerometry data was kept unfiltered to preserve low frequency components related to sleep positioning. Gravity and movement acceleration components were separated using a 2 second running average filter. All recordings were analysed in 30 second epochs aligned with sleep stage labels for feature extraction. Only epochs with concurrent sqEEG and accelerometry were kept to ensure equal modality representation across time.

 DOI 10.15203/99106-194-6-04

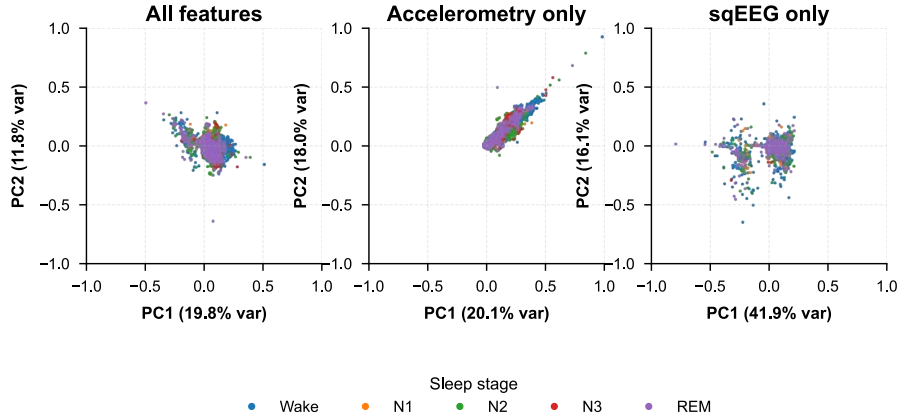


Fig. 2: Principal Component Analysis (PCA) biplots for combined (**left**), accelerometry only (**center**), and sqEEG only features (**right**). The combined modality space is primarily governed by sqEEG spectral features.

### A. *Feature Extraction*

Feature extraction was performed on each 30 second epoch, and the resulting features were combined into a multimodal table with one row per epoch. SqEEG features consisted of simple amplitude based statistics (mean, median, and variance) together with spectral band power spanning delta to gamma ranges derived using Welch's power spectral density estimates and expressed in decibels [4]. Accelerometry features included mean, median and variance of the gravity and body acceleration components for each axis. The body acceleration vector magnitude was computed using the triaxial Euclidean distance, and the same summary statistics were computed to quantify movement intensity during sleep [2].

PCA was applied to assess the redundancy between modalities and to determine whether the added modality of accelerometry contributed to the explained variance beyond what was already present in the sqEEG features. PCA was performed on three separate matrices; the combined multimodal feature set, the sqEEG only subset and the accelerometry only subset. All features were standardised prior to analysis.

In addition to PCA, other multivariate strategies could be considered to characterize both shared and unique information between the modalities. For example, Canonical Correlation Analysis (CCA) could explicitly maximize linear correlations between the two feature sets, while Partial Least Squares (PLS) could prioritize components that covary with sleep stage labels. Non-linear approaches such as kernel PCA or autoencoder based methods may reveal relationships that are not captured by linear variance based decompositions. These approaches were not pursued in order to maintain a transparent and interpretable baseline analysis.

TABLE I: Participant demographics and implantation information ($n = 20$, per protocol). Values are given in mean $\pm$ standard deviation or percentages (%).

| **Characteristic** | **Total sample** |
|---|---|
| Age (years) | $32.5 \pm 12.4$ |
| Sex (female/male) | 11/9 (55%/45%) |
| BMI (kg/m$^2$) | $24.1 \pm 4.6$ |
| Implant location (left/right) | 10/10 (50%/50%) |
| Implant orientation (temporal) | 20 (100%) |

## III. Results

The extracted features showed limited variation across sleep stage labels, with the exception of greater variation for the Wake stages. PCA was applied to quantify shared and unique variance between sqEEG and accelerometry.

The three panel PCA biplot in Fig. 2 presents the results for the combined feature set (left), the accelerometry subset (center), and the EEG subset (right). When all features were analysed together, the first two components explained 31.6% of the variance, with the dominant axis driven by EEG spectral power, particularly in the sigma, beta, and gamma bands. The second component reflected slow posture related changes from the accelerometry features, primarily mean and median values from the Y and Z axes. Most features contributed evenly to the first principal components, with no single feature dominating the explained variance.

Accelerometry alone produced a different variance structure, with 38.1% explained by the first two components and orientation based features accounting for much of the separation in that space. SqEEG alone explained 58.0% of the variance in its first two components, with high frequency spectral power dominating the first component and lower frequency power contributing to the second.

## IV. Discussion

The PCA results indicate that the chest mounted accelerometry contributes rather small unique information relative to the sqEEG in the healthy sleep cohort. Although accelerometry captures posture and gross movements, these patterns did not introduce new dominant axes of variance in the combined feature space. The gradual changes were also visible as amplitude and spectral fluctuations in the sqEEG, which reduces the amount of additional information accelerometry provides. As a result, the accelerometry contributed little orthogonal structure when analysed alongside sqEEG. Instead,

 DOI 10.15203/99106-194-6-04

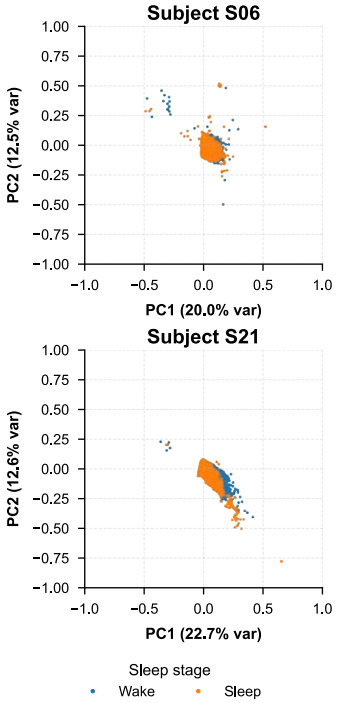


Fig. 3: PCA biplots for all features combined for subjects S06 and S21. The sleep stage labels N1 to N3, and REM have been combined for an easier view.

the multimodal PCA closely resembled the sqEEG only subset, suggesting substantial redundancy between the two modalities at the level of global variance. The broader spread of Wake epochs reflects increased movement and activity in the higher frequency sqEEG bands, but the remaining sleep stages formed overlapping clusters, highlighting the limited discriminative value of the extracted features of both modalities when reduced to simple summaries per epoch.

The limited contribution to explained variance may be partly explained by placement of the recorder in the upper chest area. Sensors worn at the wrist or near the diaphragm often capture fine limb motion, heart rate or respiratory dynamics, which are useful for sleep stage discrimination [2], [7], [8].

PCA itself captures only linear relationships and prioritises components that explain the greatest global variance. Smaller but potentially meaningful modality interactions may therefore be compressed or obscured. Furthermore, PCA performance across subjects (Fig. 3) emphasises population level variance, whereas sleep physiology and movement patterns are highly individual. Hence, personalized analyses may reveal complementary variance that does not emerge at the group level, and nonlinear analysis approaches may reveal additional structure beyond the linear variance captured by PCA. Despite these limitations, the findings provide a valuable baseline for future multimodal work.

### A. Implications for future clinical applications

While redundancy is prominent in healthy subjects, accelerometry may become more informative in clinical populations with abnormal movements or pathological sleep patterns. Conditions such as epilepsy, sleep apnoea, or parasomnias involve movement patterns that might be not already reflected in sqEEG alone [9]. In such settings, multimodal feature sets and alternative analysis strategies may provide added value for sleep stage detection. The present results therefore help contribute to the development of more targeted multimodal feature sets.

## V. Conclusion

The study assessed the relationship between sqEEG and chest mounted accelerometry in a healthy sleep cohort using a multimodal feature extraction pipeline and PCA. The combined feature space was dominated by sqEEG spectral variation, while the accelerometry contributed only minor posture related structure and added little new variance beyond what was already capture by sqEEG. These results suggest that, in healthy subjects, chest accelerometry offers limited complementary information when analysed alongside long term sqEEG. Even so, the framework developed here provides a basis for future multimodal work in clinical contexts, where movement patterns may carry greater relevance and interact more strongly with neural activity.

## Acknowledgment

The first author would like to thank UNEEG medical for data access and project supervision.

## References

[1] P. F. Viana, J. Duun-Henriksen, J. S. Winston, T. W. Kjaer, C. Skaarup, and M. P. Richardson, “Atlas of ambulatory subcutaneous electroencephalography,” *Epileptic Disorders*, vol. 25, no. 4, 8 2023.

[2] J. Razjouyan, H. Lee, S. Parthasarathy, J. Mohler, A. Sharafkhaneh, and B. Najafi, “Improving Sleep Quality Assessment Using Wearable Sensors by Including Information From Postural/Sleep Position Changes and Body Acceleration: A Comparison of Chest-Worn Sensors, Wrist Actigraphy, and Polysomnography,” *Journal of Clinical Sleep Medicine*, vol. 13, no. 11, 11 2017.

[3] I. P. Machado, A. Luísa Gomes, H. Gamboa, V. Paixão, and R. M. Costa, “Human activity data discovery from triaxial accelerometer sensor: Non-supervised learning sensitivity to feature extraction parametrization,” *Information Processing & Management*, vol. 51, no. 2, 3 2015.

[4] E. Ahrens, P. Jennum, J. Duun-Henriksen, H. W. S. Borregaard, S. S. Nielsen, N. Taptiklis, F. Cormack, B. D. Djurhuus, P. Homøe, T. W. Kjær, and M. C. Hemmsen, “The Ultra-Long-Term Sleep study: Design, rationale, data stability and user perspective,” *Journal of Sleep Research*, vol. 33, no. 6, 12 2024.

[5] “How UNEEG medical does ultra long-term EEG monitoring.” [Online]. Available: https://www.uneeg.com/data-analysis/

[6] E. Ahrens, P. Jennum, J. Duun-Henriksen, B. Djurhuus, P. Homøe, T. W. Kjær, and M. C. Hemmsen, “Automatic sleep staging based on 24/7 EEG SubQ (UNEEG medical) data displays strong agreement with polysomnography in healthy adults,” *Sleep Health*, vol. 10, no. 6, pp. –620, 12 2024.

[7] F. Ryser, S. Hanassab, O. Lambercy, E. Werth, and R. Gassert, “Respiratory analysis during sleep using a chest-worn accelerometer: A machine learning approach,” *Biomedical Signal Processing and Control*, vol. 78, 9 2022.

[8] H. Yuan, E. A. Hill, S. D. Kyle, and A. Doherty, “A systematic review of the performance of actigraphy in measuring sleep stages,” *Journal of Sleep Research*, vol. 33, no. 5, 10 2024.

[9] C. McCall and W. V. McCall, “Comparison of actigraphy with polysomnography and sleep logs in depressed insomniacs,” *Journal of Sleep Research*, vol. 21, no. 1, pp. 122–127, 2 2012.

 DOI 10.15203/99106-194-6-05

# Frequency-Based Filtering for Automated Spike Detection in Human Microneurography

Dominika Darabos, Alina Troglio, Andrea Fiebig, Anna Maxion,
Ekaterina Kutafina, Barbara Namer, Gergő Bognár, Péter Kovács

***Abstract*—Although microneurography is a powerful tool for recording activity from individual human peripheral nerve fibers, its broader adaptation is still limited due to the characteristics of the data, which prevent most existing spike detection algorithms from being used effectively. Since these are extracellular recordings collected with microelectrodes, they contain a significant amount of noise, resulting in a low signal-to-noise ratio. This noise can easily mask the spike-shaped responses of nerve fibers, whose shape can also vary over time, making it difficult to develop detectors that generalize reliably across different spike shapes. In this study, we analyze the recordings in the frequency domain and identify frequency components that can be removed with minimal compromise to the underlying neural activity. Using the weighted Hermite Variable Projection Neural Network applied to sliding-window segments of the signal, we compare classification performance on the original, low-pass filtered, and band-pass filtered data. The results identify the frequency ranges that carry the most informative components of the neural signal and demonstrate that appropriate filtering can reduce the number of false alarms to 22% of the level obtained with unfiltered data, while only modestly affecting true detections.**

***Index Terms*—microneurography, spike detection, band-pass filtering, low-pass filtering, noise reduction, variable projection, sliding-window classification, time-series analysis, VPNet**

## I. Introduction

C-type fibers [1], [2] are responsible for transmitting neural information such as nociception or pruriception within the peripheral nervous system. Understanding the patterns of these neural responses could contribute significantly to the development of treatments for neuropathic pain. When a nociceptive stimulus is detected by sensory receptors, the resulting neural activity travels through the nerve fibers toward the central nervous system. This activity can be recorded with microelectrodes inserted into the nerve bundles. The number of detectable fibers depends on on the position of the single electrode, and these fibers are referred to as the target fibers. The technique enabling such measurements is microneurography [3], [4].

Previous studies [5], [6] have shown that artificial, constant-frequency electrical stimulation evokes nearly constant response times in C-type fibers, either immediately or—depending on the stimulation frequency—after an initial period of slowing, while the introduction of additional stimuli leads to nonlinear changes in response latency. If spike-like neural responses were always cleanly observable in the signal, this temporal variability would cause less of a challenge.

In practice, microneurography recordings are extracellular and contain activity from nearby non-target fibers as well as substantial in vivo, environment-related noise, while spike shapes may vary over time. These factors further challenges the ability of detectors to generalize to a single canonical shape. Detection algorithms specifically designed for microneurography remain extremely limited. Currently, a semi-automated approach is commonly used in which experts identify evoked spikes based on expected latencies using a vertical waterfall-style visualization [7]. While spikes may be partially masked by noise, vertical alignment allows reliable identification with stable conduction velocity. However, this approach provides only partial detection, as spontaneous spikes cannot be captured. Once spikes are detected, spike-sorting methods are applied to assign events to individual fibers, for which algorithms already exist [8].

In our previous work [9], we presented, to our knowledge, the first neural network evaluated on microneurography data for spike detection. The sequential recordings were converted into sliding-window segments, and a knowledge-augmented classification model—the weighted Hermite Variable Projections Neural Network (WHVPNet) [10]—was used to predict whether a given window contained spike-like activity. That work focused primarily on handling extreme class imbalance and extracting shape-based features for spike characterization.

In the present study, we shift attention to the signal structure itself. We analyze the raw recordings in the frequency domain to identify potential patterns related to the underlying neural activity and to determine which frequency bands are dominated by noise. These findings are then used to create low-pass and band-pass filters designed to remove uninformative frequency components with minimal distortion of the neural signal. The filtered recordings are integrated into the same WHVPNet framework, allowing us to directly compare model performance on raw versus frequency-filtered data.

## II. Methods

All training, testing, and evaluation were performed on a single in vivo human microneurography recording containing two C-type fibers [6]. The recording was sampled at 10 kHz, and the total duration of the recording is approximately 13 minutes. The study was approved by the Ethics Board of the University Hospital RWTH Aachen (Vo-Nr. EK141-19), and written informed consent was obtained. Ground-truth spike labels were generated by a single expert using

D.D., G.B., and P.K. are with the Eötvös Loránd University, Budapest, Hungary. A.T., A.F., and A.M are with the RWTH Aachen University, Aachen, Germany. E.K. is with the University of Cologne, Cologne, Germany. B.N. is with the University of Würzburg, Würzburg, Germany. The project was supported by the University Excellence Fund of Eötvös Loránd University, Budapest, Hungary (ELTE). P. Kovács was supported by the János Bolyai Research Scholarship of the Hungarian Academy of Sciences.

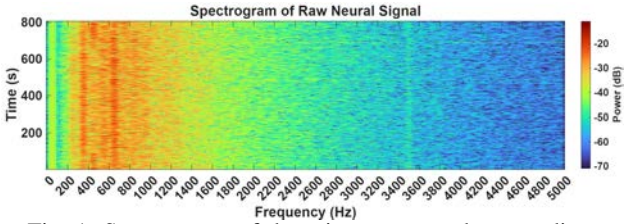


Fig. 1: Spectrogram of the microneurography recording, showing the time–frequency representation.

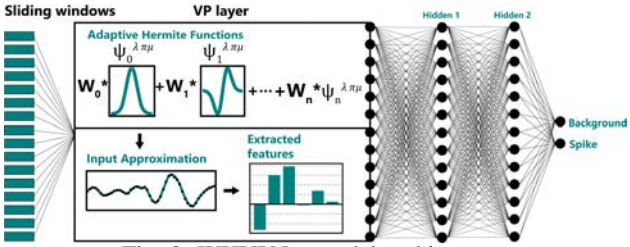


Fig. 3: WHVPNet model architecture.

the vertical-alignment procedure described in [6]. The signal underwent fixed hardware filtering during recording, however, as the hardware predates detailed documentation, precise filter specifications are unavailable, and the data were treated as raw signals. Due to high inter-recording variability in microneurography, generalization to other recordings is left to future work involving calibration and transfer-learning strategies.

### A. Frequency-domain analysis

To characterize the spectral structure of the recordings, short-time Fourier transforms were computed using a 256-sample window, 200-sample overlap, and a 512-point fast Fourier transformation at a 10 kHz sampling rate. The resulting spectrogram (Fig. 1) indicated that most neural energy resides in lower-frequency bands, which motivated the subsequent filtering strategy.

### B. Filtering

Guided by the spectral analysis, 4th-order Butterworth low-pass and band-pass filters were applied (Fig. 2) using zero-phase forward–reverse filtering. Several band-pass configurations with varying lower and upper cutoff frequencies were evaluated. Because the relevant neural bands are tightly clustered, a single band-pass setup was used rather than a multiband combination.

### C. Integration with VPNet

Filtered data were processed using the same sliding-window WHVPNet framework (Fig. 3) as in the previous work. The model adapts Hermite basis functions through learnable parameters to capture spike-shape variability. Their combination provides an input approximation from which features are extracted, followed by fully connected layers that perform binary, window-level spike classification, deciding between spike and non-spike. The number of extracted features and Hermite-system weights are model parameters, selected through prior hyperparameter optimization. Due to the model's compact size (64 parameters), two evaluation strategies were considered:

- retraining VPNet on each filtered dataset, and
- applying filters only during validation/testing while keeping the original model unchanged.

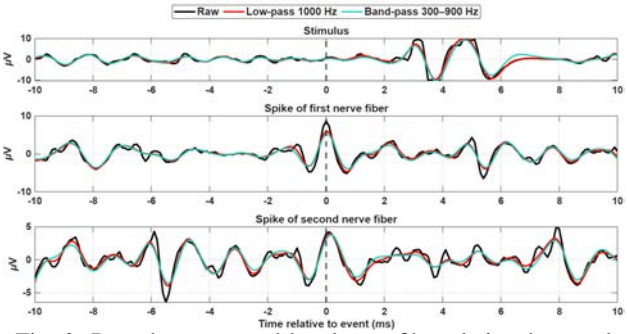


Fig. 2: Raw, low-pass and band-pass filtered signal around the a stimulus, and two types of nerve fiber responses.

### D. Metrics

Using overlapping-window preprocessing, metrics defined in [9] were applied. Each window was initially treated as an independent prediction in a *window-wise* manner. Consecutive positive windows corresponding to the same spike were then *merged*, with the *proximity* metric removing merged predictions of only one window. The *latency* metric further ignored predictions inconsistent with expected response timing.

Given the extreme class imbalance, recall was used to evaluate detection performance, while false positives were quantified as the ratio per true positive (FP/TP).

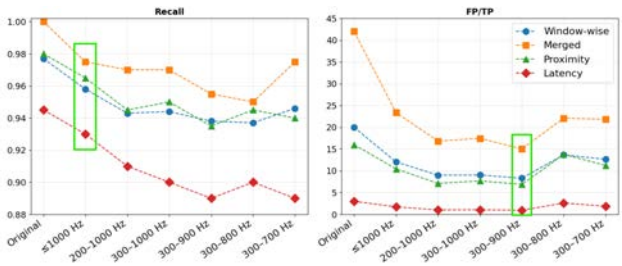


Fig. 4: Recall and FP/TP for different filtered signals using the introduced metrics, evaluated on the validation set after retraining. Green rectangles mark the best setting per subplot.

## III. Results

The spectrogram (Fig. 1) shows that most neural energy is concentrated below 1000 Hz, with prominent bands between 300–700 Hz. A higher-frequency band around 3500 Hz was also visible, but including it in a multiband filter had no measurable impact on performance. Based on the 0–1000 Hz range, Fig. 4 compares model performance across different filtered signals using the metrics described in Sec. II-D. Among the filtered signals, the low-pass filter achieved the highest recall across all metrics, while the 300–900 Hz band-pass filter minimized the FP/TP ratio.

TABLE I: Performance metrics for the datasets using the original signal, the low-pass filtered (1000 Hz), and a band-pass filtered signal (300–900 Hz). Results are shown for retrained models (R) and for the baseline model trained only on the original data but evaluated on the filtered signals (E). An 80% decision threshold is used.

| Dataset | Condition | Window-wise | Merged Windows | Proximity | Proximity + Latency |
|---|---|---|---|---|---|
| **RECALL** | | | | | |
| Validation | **Original (R)** | **0.977** | **1.000** | **0.980** | **0.945** |
| Validation | Low-pass (R) | 0.958 | 0.975 | 0.965 | 0.930 |
| | Low-pass (E) | 0.96 | 0.98 | 0.96 | 0.925 |
| Validation | Band-pass (R) | 0.938 | 0.955 | 0.935 | 0.890 |
| | Band-pass (E) | 0.945 | 0.97 | 0.95 | 0.9 |
| Test | **Original (R)** | **0.957** | **0.983** | **0.967** | **0.943** |
| Test | Low-pass (R) | 0.916 | 0.94 | 0.934 | 0.918 |
| | Low-pass (E) | 0.912 | 0.935 | 0.918 | 0.89 |
| Test | Band-pass (R) | 0.884 | 0.92 | 0.894 | 0.853 |
| | Band-pass (E) | 0.888 | 0.92 | 0.89 | 0.85 |
| **FP / TP** | | | | | |
| Validation | Original (R) | 20 | 42 | 16 | 3 |
| Validation | Low-pass (R) | 12.04 | 23.37 | 10.45 | 1.72 |
| | Low-pass (E) | 11.3 | 22.0 | 8.76 | 1.46 |
| Validation | Band-pass (R) | 8.31 | 15.02 | 6.88 | 0.94 |
| | Band-pass (E) | **8.15** | **14.94** | **6.37** | **0.81** |
| Test | **Original (R)** | **24** | **51** | **18** | **3** |
| Test | Low-pass (R) | 13.17 | 25.44 | 11.08 | 1.42 |
| | Low-pass (E) | 11.78 | 23.37 | 8.66 | 1.21 |
| Test | Band-pass (R) | 9.47 | **16.33** | 7.51 | 0.84 |
| | Band-pass (E) | **9.21** | 16.57 | **6.96** | **0.68** |

This illustrates the deep-rooted trade-off between recall and FP/TP: improving one metric tends to compromise the other. Therefore, both filters were selected for further evaluation.

As summarized in Table I, retraining the model on filtered data (R) generally produced similar outcomes to evaluating the original model on filtered data only (E), with some cases performing better in the latter. While filtering caused a moderate decrease in recall, the FP/TP ratio dropped dramatically—by roughly 45% to 80%—demonstrating that the reduction in false positives far outweighs the loss in sensitivity.

Performance on the test set was slightly lower than on the validation set, which is expected given natural variability and the limited size of in vivo microneurography recordings. Despite this, the overall trends and improvements observed with filtering remain consistent.

## IV. Conclusion

In this work, the frequency content of noisy microneurography recordings was analyzed, revealing that the majority of neural-related information is located below 1000 Hz, despite power being present up to 5000 Hz. Applying a low-pass filter at 1000 Hz removed approximately 80% of the spectral power while preserving the relevant neural signal. A selected band-pass filter, targeting 300–900 Hz, excluded around 88% of the total frequency content, further isolating the frequency range most informative for spike detection. The primary benefit of these filtering steps was a substantial reduction in false positives. Although to some extent, recall decreased as well, this reduction was relatively small compared to the improvement in false positives. In some cases, the FP/TP ratio dropped to just 22% of the value obtained with the unfiltered data. As the data is inherently filtered during recording, supposedly, re-annotating the signal would not change the spike occurrences. Future work could extend this approach by incorporating learnable filter parameters, enabling the pipeline to automatically identify cutoff frequencies that optimize model performance.

## References


[1] C. K. Arcilla and P. Tadi, *Neuroanatomy, Unmyelinated Nerve Fibers*. Treasure Island, FL: StatPearls Publishing, 2023.

[2] B. Namer and A. Lampert, "Functional signatures of human somatosensory c fibers by microneurography," *Pain*, vol. 166, no. 10, pp. 2220–2235, 10 2025.

[3] A. B. Vallbo and K.-E. Hagbarth, "Activity from skin mechanoreceptors recorded percutaneously in awake human subjects," *Experimental Neurology*, vol. 21, no. 3, pp. 270–289, 1968.

[4] H. E. Torebjörk and R. G. Hallin, "Identification of afferent C units in intact human skin nerves," *Brain Research*, vol. 67, no. 3, pp. 387–403, 1974.

[5] E. Kutafina, S. Becker, and B. Namer, "Measuring pain and nociception: Through the glasses of a computational scientist. transdisciplinary overview of methods," *Frontiers in Network Physiology*, vol. 3, 02 2023.

[6] E. Kutafina, A. Troglio, R. de Col, R. Röhrig, P. Rossmanith, and B. Namer, "Decoding neuropathic pain: Can we predict fluctuations of propagation speed in stimulated peripheral nerve?" *Frontiers in Computational Neuroscience*, vol. 16, 2022.

[7] M. Schmelz, C. Forster, R. Schmidt, M. Ringkamp, H. O. Handwerker, and H. E. Torebjörk, "Delayed responses to electrical stimuli reflect C-fiber responsiveness in human microneurography," *Experimental Brain Research*, vol. 104, no. 2, pp. 331–336, 1995.

[8] A. Troglio, P. Konradi, A. Fiebig, A. Pérez Garriga, R. Röhrig, J. Dunham, E. Kutafina, and B. Namer, "Supervised spike sorting feasibility of noisy single-electrode extracellular recordings: Systematic study of human c-nociceptors recorded via microneurography," *PLOS ONE*, vol. 20, no. 9, pp. 1–28, 09 2025.

[9] D. Darabos, A. Troglio, A. Fiebig, A. Maxion, E. Kutafina, B. Namer, G. Bognár, and P. Kovács, "Explainable spike detection algorithm on c-fiber microneurography data using weighted hermite variable projection neural networks," in *Proc. 47th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, July 2025, accepted.

[10] T. Dozsa, C. Bock, J. Meier, and P. Kovács, "Weighted Hermite variable projection networks for classifying visually evoked potentials," *IEEE Transactions on Neural Networks and Learning Systems*, October 2024.

 DOI 10.15203/99106-194-6-06

# Hardware Prototype of a Wearable and End-to-End EEG-Gaming Controller using Flower Electrodes

Nick Skillandat*, Hasan Nehir*, Christopher Ringhofer, Lukas Einhaus,
Gregor Schiele, Patrique Fiedler, Andreas Erbslöh

***Abstract*—This paper presents the front-end performance of a hardware prototype to enable an EEG-based gaming controller as wearable brain-computer interface (BCI). The system achieves high signal fidelity through an eight channel analog front-end (AFE) with active-shielded cables, high input impedance ($>1\,\mathrm{G\Omega}$), ultra-low noise design ($390.96\,\mathrm{nV_{rms}}$, $28.93\,\mathrm{nV}/\sqrt{\mathrm{Hz}}$@$10\,\mathrm{Hz}$) and high suppression of harmonic distortions ($114.1\,\mathrm{dB}$). Measurement results show a second-order band-pass filter ($0.52\,\mathrm{Hz}$–$543.1\,\mathrm{Hz}$) with a total gain of 37.34 dB. Active shielding significantly suppresses 50 Hz power-line interference for clean signal acquisition, while floating cables introduce strong artifacts. The prototype demonstrates robust, low-noise EEG recording suitable for (i) BCI applications with enough computational resources to deploy a task-related data stream pre-processing and embedded deep learning techniques on hardwareand (ii) using flower electrodes to increase user-friendliness.**

***Index Terms*—brain-computer interfaces, electroencephalography, embedded systems, flower electrodes, wearable electronic**

## I. Introduction

Brain-computer interfaces (BCIs) translate neural activity into control signals without relying on peripheral nerves or muscles. This opens up new possibilities in neurorehabilitation and enables people with sensory and motor impairments to regain their abilities to master everyday tasks [1].

Electroencephalography (EEG) is a well established, non-invasive and low cost way to record brain activity, especially when compared to other methods, such as invasive recording device like neural implants using Utah array [2]. Conventional EEG electrodes need a conductive gel to function correctly. This is an additional hurdle for using such solutions in everyday life. The recently developed so called flower electrodes avoid this problem and achieve a comfortable reliable fit without conductive gel [3]. These developments are crucial for bringing BCIs out of the laboratory and into everyday life [4].

We introduce a hardware prototype to enable an EEG-based gaming controller. This prototype is able to record external signals with high artefact suppression and with low-noise which allows to use the flower electrodes for sensing neural activity. Also, enough computational resources are available for an embedded pre-processing and machine learning in order to recognise the movement intention to control a video game avatar from the EEG signal in real-time. In related work, Martisius et al. [5] show an implementation to detect three different actions with an accuracy of $80\,\%$. Newer deep learning models, like AMEEGnet, achieve a precision up to $95.49\,\%$ for motor imagery EEG decoding [6].

*N. Skillandat and H. Nehir are students (Master, Cyber Physical Systems) at the University of Duisburg-Essen and sharing the first authorship through equal contribution.

P. Fiedler is with the Department of Data Analysis in Life Sciences, Technical University Illmenau, Germany.

C. Ringhofer, L. Einhaus, G. Schiele, A. Erbslöh are with the Intelligent Embedded Systems Lab, University of Duisburg-Essen, Germany.

## II. Hardware Design

This section describes (i) the design constraints for the hardware development of the EEG-based Gaming Controller using dry-flower electrodes, (ii) the hardware concept, and (ii) the methods for defining the hardware metrics.

### A. Design Constraints

To enable flower electrodes in EEG setups, the input impedance of the input amplifier stage needs to be higher than $1\,\mathrm{G\Omega}$ [3]. The wires should be actively shielded to eliminate parasitic capacities. This shielding technique reduces the impact of artifact coupling effects from the environment, e.g $50\,\mathrm{Hz}$ power line. To recognize movement intentions reliably, the analog front-end (AFE) should be designed with ultra-low noise techniques to sense activities in the lower µV-range ($<2\,\mu\mathrm{V_{pp}}$, $<1\,\mu\mathrm{V_{rms}}$). To handle a recording setup with channel number, the used analog-digital converter (ADC) should sample all input channels simultaneously (24-bit, $\leq 1\,\mathrm{kHz}$). As the solution needs to be wearable, the AFE should be low-power to facilitate long recording sessions.

For improved privacy and security, the data pre-preprocessing and machine learning (ML) model should run on an embedded system, e.g., using hardware accelerators on a field-programmable gate-array (FPGA). Regarding real-time processing, the movement intention should be read-out with a latency below $300\,\mathrm{ms}$ [5].

### B. Hardware Concept

Figure 1 (top) shows the hardware concept to fulfill the design constraints. It depicts the interaction between the different modules and a detailed description for one AFE channel.

Each AFE channel consists of three stages. The first stage includes the selection of the reference signal (system ground or reference electrode) and the amplifiers to actively shield the SMA cables of recording and reference electrodes.

The high-Z input amplifier in the second stage is an instrumentation amplifier (AD8422, Analog Devices). It provides high input impedance ($>200\,\mathrm{G\Omega}$) in common mode and its gain is configurable (10-200 V/V) using a digital potentiometer. A Sallen-Key circuit enables a second-order band-pass

 DOI 10.15203/99106-194-6-06

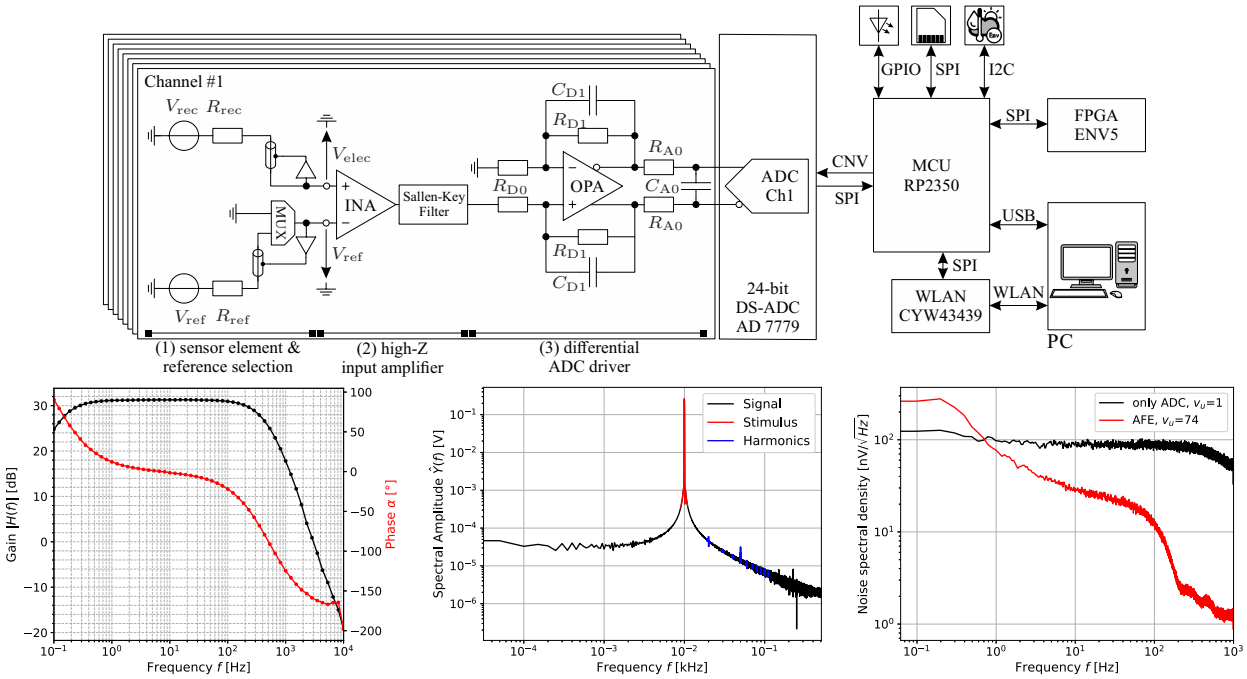


Fig. 1: Top: Hardware concept for sensing and processing EEG signals including a detailed circuit schematic of one AFE channel, Down: (left) results from the frequency response analysis from AFE input to ADC input, (middle) spectrum of the ADC output for extracting the harmonic distortion, (right) input noise spectral density of the AFE.

filter characteristic. The third stage includes a 24-bit delta-sigma ADC (AD7779, Analog Devices) with pre-amplifier in fully-differential configuration and an optional low-pass filter.

A microcontroller (RP2350A, Raspberry Pi) is responsible for controlling the data acquisition and configuring all devices. The data is acquired with configurable sampling rates. The MCU hosts a pipeline to process digital signals in real-time, e.g. downsampling, filtering, window sequencing and deep learning. Though, the current implementation only supports sending a raw ADC data stream. An acceleration sensor (BMI270, Bosch) and a temperature sensor (SHT41, Sensiron) provide additional context to interpret the EEG signals.

The ADC raw and sensor data can be stored on a microSD card or transmitted to an external computer for further offline and online processing via USB or WLAN. Alternatively it can be passed to a custom FPGA board (ENV5), specialised for embedded deep learning inference. Figure 2 shows the assembled hardware on the left and a zoom on one AFE channel highlighting the three stages. This setup can collect EEG signals of up to 8 channels simultanouesly. The electrodes are connected via an sensor board and active-shielded SMA cables. Also, a reference electrode and a ground electrode can be connected. The power supply is delivered from the USB header by connecting a power bank to it.

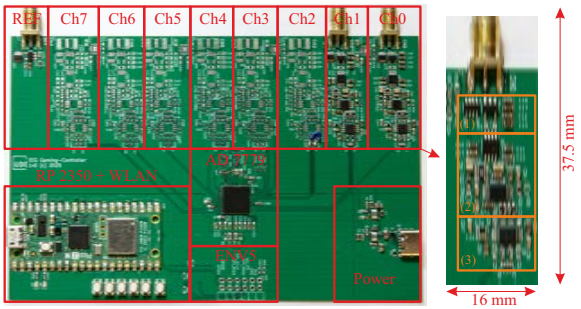


Fig. 2: Picture of the manufactured and assembled hardware setup used as EEG-based gaming controller with active-shielded flower electrodes (size: 140 mm x 90 mm).

### C. Characterization Methods

To characterize the performance of the AFE, we explain the procedure for measuring the hardware properties: (i) gain and filter characteristic, (ii) harmonic distortion and (iii) noise.

*1) Gain and filter characteristic:* Here, a frequency response analysis (FRA) is done by using the oscilloscope R&S MXO44. During the measurement, a sinusoidal waveform ($20\,\mathrm{mV_{pp}}$) is applied to the AFE input and its frequency is sweeped from 0.1 Hz to 10 kHz. To extract the gain profile, the positive output of the differential ADC driver is fed back into the FRA in order to calculate the transfer function.

*2) Harmonic Distortion:* We measure the transients by capturing the ADC data stream for 600 s at a sampling rate of 2 kHz and applying a sinusoidal signal ($20\,\mathrm{mV_{pp}}$, 10 Hz) to the AFE input, which is generated with the R&S MXO44. To determine the non-linear amplification effect, the spectrum is determined using the fast-fourier transformation (FFT) and the amplitudes of the harmonics and of the stimulus are used to calculate the total harmonic distortion ($THD$).

*3) Noise properties:* A transient measurement is performed by capturing the ADC data stream for 300 s at a sampling rate of 2 kHz and shorting the AFE input to system ground. By determining the power spectrum density (PSD) from the data stream, the noise spectral density can be directly extracted

and the effective input noise is calculated by integrating the spectrum for given bandwidhts and dividing it by the gain. The effective input noise is crucial for recording weak EEG signals with amplitudes in the lower µV-range.

## III. Results

This section presents the measured metrics on real hardware.

Figure 1 (down, left) shows the FRA results. Due to the Sallen-Key filter, all AFE channels have a second-order band-pass filter characteristic with a single-side gain of $31.34\,\mathrm{dB}\pm0.02\,\mathrm{dB}$ ($\equiv 37\,\mathrm{V/V}$), a high-pass corner frequency of $0.52\,\mathrm{Hz}\pm0.09\,\mathrm{Hz}$ and a low-pass corner frequency of $543.1\,\mathrm{Hz}\pm0.95\,\mathrm{Hz}$.

Figure 1 (down, middle) shows the spectrum for calculating $THD$. This metric is calculated with the ratio between the squared sum of the harmonic amplitude (highlighted in blue) and stimulus (in red) logarithmically. The proposed AFE achieves a $THD$ of $114.1\,\mathrm{dB}\pm2.01\,\mathrm{dB}$.

Figure 1 (down, right) shows the results of the input noise spectral density for two cases: (i) shorting the ADCs input to system ground for determining the ADC properties and (ii) shorting the AFEs input to system ground for determinig the AFE properties. In the first case, the ADC has a constant noise density of $89\,\mathrm{nV}/\sqrt{\mathrm{Hz}}$. Compared to the second case, the noise spectrum density includes pink noise at frequencies lower than $2.4\,\mathrm{Hz}$ and thermal noise frequencies higher than $28.93\,\mathrm{nV}/\sqrt{\mathrm{Hz}}$. Due to the filter characteristic, the noise density drops at frequencies higher than $300\,\mathrm{Hz}$. Integrating over the total bandwidth at a total gain of $37.34\,\mathrm{dB}$ ($\equiv$ $74\,\mathrm{V/V}$), the ADC channels have an effective input noise voltage of $2.31\,\mu\mathrm{V_{rms}}\pm0.22\,\mu\mathrm{V}$ and one AFE channel achieves $390.96\,\mathrm{nV_{rms}}\pm20.32\,\mathrm{nV}$. From transient, a peak-to-peak value below $1.8\,\mu\mathrm{V_{pp}}$ is observed. Commercial systems have similar values, like g.tec g.Nautilus ($<600\,\mathrm{nV_{rms}}$) and ANT Neuro eego 8 ($<1\,\mu\mathrm{V_{rms}}$).

Compared to the design constraints, this prototype fulfills all requirements. The hardware consumpes $48.5\,\mathrm{mW@5\,V}$ during data acquisition for two assembled AFE channels.

## IV. Discussion

In this section, we discuss the impact of the shielding technique on the data acquisition. We emulate EEG signals with a sinusoidal waveform ($20\,\mathrm{mV_{pp}}$, $10\,\mathrm{Hz}$) which is applied on the sensor board for two cases: active shielded and floating electrode. System ground is connected over a two-meter SMA cable as a reference signal. Figure 3 (top) shows the recording results with the active-shielded cables. We measure the ADC data stream, the applied electrode signal to the sensor board and the shield signal on the sensor board. The outcome is that all signals are perfectly in-line and in the spectrum, only the signal of interest is available. Applying floated shielding, the output of the feedback amplifier is disconnected from outer conductor of the SMA cable. Figure 3 (bottom) shows the response directly. Opposed to active shielding, floated shielding results in powerline interference with amplitudes similar to those of the input signal. This leads to higher effort in the rest of the processing pipeline for removing these artefacts using high-order notch filter.

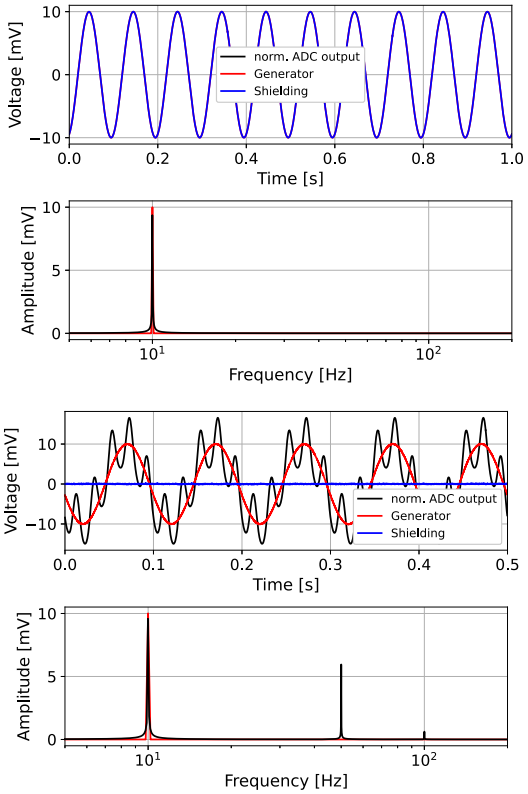


Fig. 3: Experimental results (transient signal and spectrum) for investigating the impact of shielding techniques on the data acquisition (Top: Active, bottom: Floated shielding).

## V. Conclusion

This paper presents a hardware prototype for sensing EEG signals for up to eight channels with active shielded cables in order to use flower electrodes. The analog front-end has a second-order band-pass filter ($0.52\,\mathrm{Hz}$–$543.1\,\mathrm{Hz}$) with configurable gain. Applying a total gain of $37.34\,\mathrm{dB}$ ($\equiv 74\,\mathrm{V/V}$), it has a total harmonic distortion of $114.1\,\mathrm{dB}$ and an input noise of $390.96\,\mathrm{nV_{rms}}$ at a sampling rate of $2\,\mathrm{kHz}$.

The next steps are to integrate the hardware and flower electrodes into a wearable cap. This will be used for investigating the end-to-end signal processing pipeline to enable an EEG-based Gaming Controller with pre-processing and embedded deep learning techniques. Also, online training of the deep learning architecture during runtime should be investigated for optimizing a basic model to each patient.

## References

[1] X. Gu et al., "EEG-Based Brain-Computer Interfaces (BCIs): A Survey of Recent Studies on Signal Sensing Technologies and Computational Intelligence Approaches and Their Applications," IEEE/ACM Trans. on Comp. Biology and Bioinformatics, vol. 18, no. 5, pp. 1645–1666, 2021.

[2] L. Buron et al., "Technical survey of end-to-end signal processing in BCIs using invasive MEAs," IOP Jour. Neu. Eng., vol. 21, no. 5, 2024.

[3] I. F. Warsito et al., "Flower electrodes for comfortable dry electroencephalography," Scientific Reports, vol. 13, no. 1, p. 16589, 2023.

[4] A. Saibene et al., "EEG-Based BCIs on Motor Imagery Paradigm Using Wearable Technologies: A Review," Sensors, vol. 23, no. 5, 2023.

[5] I. Martišius et al., "A Prototype SSVEP Based Real Time BCI Gaming System," Comput, Intelli. and Neuroscience, vol. 2016, p. 3861425, 2016.

[6] X. Wu et al., "AMEEGNet: attention-based multiscale EEGNet for effective motor imagery EEG decoding," Fr. Neurorobot., vol. 19, 2025.

 DOI 10.15203/99106-194-6-07

# Benchmarking ECG R-peak Detectors in Offline and Real-time Settings across Heterogeneous Datasets

S. Costantini, F.A. Storm, A.M. Bianchi

***Abstract*—Reliable R-peak detection is essential for effective ECG analysis. However, standard benchmarking procedures often rely on noise-free datasets, such as the MIT-BIH Database, potentially overestimating performance in real-world scenarios. This study seeks to address this limitation by evaluating ten open source R-peak detection algorithms across five heterogeneous datasets, which include daily-living conditions, different levels of physical activity, and normal and arrhythmic rhythms. Offline benchmarking identified Neurokit, Kalidas, and Rodrigues as the top performers; however, cross-dataset analysis revealed Neurokit as the most robust algorithm, maintaining high stability even under severe motion artifacts and irregular ECG morphologies. In online benchmarking, Neurokit proved its superior performance, combining high accuracy, low latency, and high computational efficiency. A 3-second observation window was identified as the ideal operating point for Neurokit, achieving an optimal Jitter F1 estimated marginal mean of 0.93 across all datasets.**

***Index Terms*—ECG, R-peak detection, Benchmark, Online**

## I. Introduction

Reliable R-peak detection is crucial for biomedical applications like heart rate variability analysis, affective computing, and closed-loop human-computer interaction. Over the past decades, many algorithms for R-peak detection have been proposed, ranging from classical approaches such as those implemented by Pan and Tompkins [1] or Hamilton [2], to more recent adaptive or wavelet-based algorithms [3], [4], [5]. Despite the open source availability of these algorithms, benchmarking studies that provide systematic statistical comparisons across several R-peak detection algorithms and different open access datasets remain limited. In fact, most benchmarking studies on R-peak detection algorithms rely almost exclusively on the MIT-BIH Arrhythmia Database [6], whose noise-free recordings tend to overestimate algorithmic performance. To address these limitations, Porr et al. [7] introduced the Glasgow University Database (GUDB), a novel ECG dataset including daily life-related conditions and a more rigorous performance metric, Jitter F1, accounting both for temporal jitter and detection accuracy. On these bases, the present work had two main objectives. First, the offline benchmark study by Porr et al. [7] was extended by broadening the benchmark analysis from one to five heterogeneous, publicly available ECG datasets. Second, R-peak detectors performance and computational efficiency were also investigated through an online benchmark study by simulating real-time data streaming using different observation window lengths.

S. Costantini is with the Department of Electronics, Information and Bioengineering, Politecnico di Milano, Italy, e-mail: simone.costantini@polimi.it

F.A. Storm is with the Scientific Institute, IRCCS "E.Medea", Bosisio Parini, Italy

A.M. Bianchi is with the Department of Electronics, Information and Bioengineering, Politecnico di Milano, Italy

## II. Methods

### A. R-peak detectors

Ten extensively used R-peak detection algorithms were tested: Pan-Tompkins [1], Hamilton [2], Martinez [8], Christov [9], Gamboa [10], Elgendi [4], Engelsi [11], Kalidas [3], the Python-based unpublished Neurokit algorithm for R-peak detection [12], and Rodrigues [5]. The main inclusion criteria were the availability of the R-peak detectors as open source Python codes and their suitability to be used in real-time.

### B. ECG datasets

Five open access ECG datasets were selected for both offline and online benchmark studies: (i) the MIT-BIH Arrhythmia Database [6], containing records from 47 subjects and focusing on sinus rhythm or arrhythmia; (ii) the GUDB [13], including ECGs from 25 subjects performing 5 daily-life physical tasks (sitting, walking, jogging, hand-biking, and doing maths); (iii) the Fantasia Database [14], reporting records from 40 healthy subjects while watching a movie; (iv) the Lobachevsky University Database (LUDB) [15], which collects 200 ten-second ECG records from subjects with and without arrhythmia; and (v) the High Intensity Exercise (HIE) dataset [16], containing ECGs from 20 subjects performing a 5-phases maximal exercise test on a cycle ergometer. These datasets were selected for their variability in the study population, experimental conditions, and signal quality, allowing a wide and heterogeneous evaluation of the performance of the R-peak detectors. Each dataset was provided with manual R-peak annotations performed by experts, which served as the reference in the benchmark studies.

### C. Benchmark studies

Benchmark studies were performed using *Python 3.10*.

*1) Offline benchmark:* In the offline assessment, each R-peak detection algorithm was applied to the full length of the raw ECG recordings from all selected datasets and subjects, thus simulating a batch processing scenario where the entire ECG signal is available at once.

*2) Online benchmark:* In the online assessment, real-time data streaming was simulated to evaluate the performance and computational efficiency of the R-peak detectors. For each dataset, condition and subject, the related ECG signal was sequentially processed using a sliding window of fixed length T, with T ranging between 1, 2, 3, 4, and 5 seconds, and an overlap percentage of 20%. To ensure signal continuity across windows, the R-peak detection algorithm was applied only to the first 90% of the window samples, while the remaining

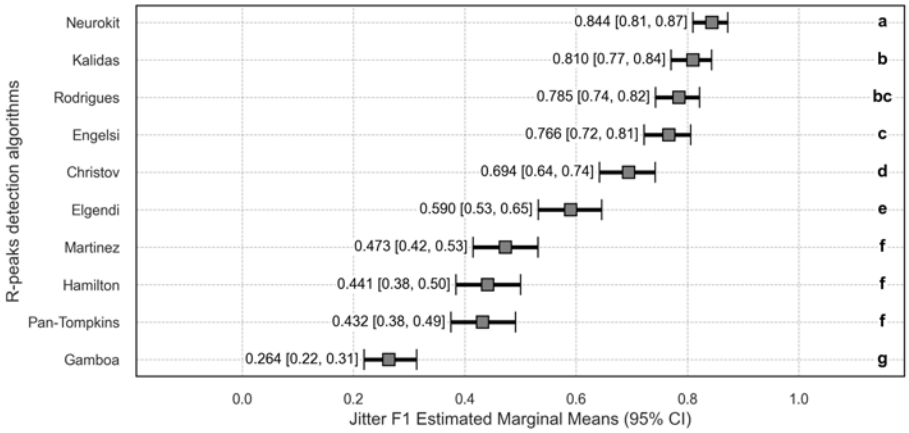


Fig. 1: Estimated Marginal Means and 95% Confidence Intervals for Jitter F1 across all ten R-peak detection algorithms in the offline benchmark. Results of post-hoc comparisons are reported using the Compact Letter Display notation.

10% were included in the subsequent, overlapping window. To avoid repeated detections due to overlapping regions, peaks in the current window were compared to those in the preceding one, and any peak within a 250 ms temporal tolerance of a previously detected event was rejected as a duplicate.

### D. Evaluation metrics

The Jitter F1 (JF) metric [7] was computed for the offline benchmark. JF was defined as the product of the F1 score, where a detected peak is valid if falling within $\pm 50$ ms of the reference peak [17], and a penalty term for temporal dispersion. In the online benchmark, CPUTime, along with JF, was computed as the average processing time needed to run the R-peak detector algorithm on a single data chunk.

### E. Statistical analysis

All statistical analyses were done using R (R 4.4.1, Vienna, Austria) within the RStudio environment. Generalized Linear Mixed Models (GLMM) were used to perform statistical analyses on the results of both benchmark studies. As for offline benchmark, the detection algorithm was included as fixed effect, while the datasets and subjects were treated as random effects to account for the hierarchical structure of the data. For online benchmark, the observation window length was included as an additional fixed effect alongside the detection method. Random effects were consistent with the offline model. GLMMs were fitted separately for JF and CPUTime metrics, although different distributions were chosen: a beta distribution with Smithson-Verkuilen correction [18] was used for JF, while CPUTime was fitted using a gamma distribution. For all statistically significant effects ($p < 0.05$), post-hoc pairwise comparisons were performed using Estimated Marginal Means (EMMs). Holm correction was applied to adjust for multiple testing.

## III. Results and Discussion

### A. Offline benchmark

The GLMM analysis revealed a significant effect of the detection algorithm on JF ($p < 0.001$), as well as a remarkable goodness-of-fit (conditional $R^2$: 0.84). Fig. 1 shows the JF EMMs and their corresponding 95% confidence intervals. Post-hoc comparisons revealed performance clusters among detectors: Neurokit was the top-performing algorithm, reporting the highest JF EMM and significantly outperforming all other detectors. This was likely due to its algorithm based on the steepness of the absolute derivative, which increases robustness against baseline wander compared to amplitude-based thresholds methods. It was followed by Kalidas, while Rodrigues performed comparably to Kalidas and Engelsi. As a result, Neurokit, Kalidas, and Rodrigues were selected as the top three detectors for the subsequent analyses. A generalization study was also performed to assess the robustness of the top-scoring detectors by examining the JF distribution across datasets and experimental conditions. All three detectors exhibited near-optimal performance across the majority of GUDB conditions involving low-to-moderate physical activity (i.e., sitting, maths, walking), as well as in the Fantasia dataset and the HIE phases 1, 2, and 5, consistently reaching median JF scores $> 0.99$ with minimal dispersion (IQR $< 0.1$). However, performance deterioration became clear in high-motion scenarios. Under the GUDB jogging condition, Neurokit reported high stability (median JF: 0.99, IQR: 0.01), whereas Rodrigues JF drop to 0.93, and Kalidas JF to 0.71. Similar trends were observed for the HIE phases 3 and 4 (i.e., high-intensity activity phases). In the MIT-Arrhythmia dataset, Neurokit reached a median JF of 0.99 (IQR: 0.04), consistently outperforming both Kalidas (JF: 0.78) and Rodrigues (JF: 0.78), which showed greater variability (IQR $> 0.22$).

### B. Online benchmark

Following the results of the offline benchmark, the online one was conducted exclusively on the top three performing R-peak detectors (i.e., Neurokit, Kalidas and Rodrigues) across the selected datasets. GLMMs fitted on JF and CPUTime reported an excellent goodness-of fit (conditional $R^2$: 0.98 for JF, 0.99 for CPUTime). Figure 2 reports the EMMs and 95% confidence intervals of the three top-scoring algorithms across

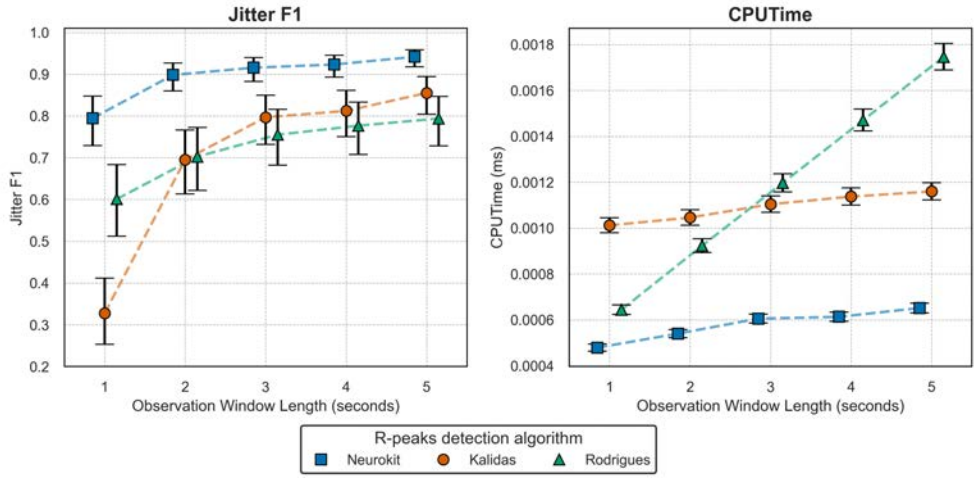


Fig. 2: Estimated marginal means and 95% confidence intervals for Jitter F1 (left panel) and CPUTime (right panel) for the top three R-peak detection algorithms across different observation window lengths (T=1s to 5s) in the online benchmark.

different observation window lengths for both metrics. The GLMM analysis revealed a significant effect of the detection algorithm, the observation window length, and their interaction on JF ($p < 0.001$). Neurokit consistently obtained the highest JF across all window lengths, proving its excellent robustness even with very short observations. Additionally, for both Neurokit and Kalidas, the increase in JF was statistically significant only up to $T = 3s$, after which performance plateaued ($p > 0.05$). Conversely, Rodrigues needed longer windows to maximize its performance ($p < 0.001$), thus suggesting that Neurokit and Kalidas may be better suited for applications with low latency, as they achieve optimal performance with shorter windows compared to Rodrigues. Regarding computational cost, Neurokit was significantly faster in absolute terms ($p < 0.001$), though exhibiting a similar slope to Kalidas, whereas Rodrigues showed a steeper linear increase in CPUTime ($p < 0.001$). To sum up, Neurokit represents the optimal choice for real-time applications, providing the highest accuracy and temporal precision at the lowest computational cost. Also, results may suggest an optimal operating point at a window length of $T = 3s$, where Neurokit maximized its JF score while maintaining negligible processing time. While dependent on sampling rate (typically $> 250$ Hz in these datasets), a 3-second window offers a balanced trade-off for most standard wearable applications.

## IV. Conclusion

The present work addressed the limitations of standard ECG benchmark studies by evaluating ten open source R-peak detection algorithms across five heterogeneous datasets. The Jitter F1 performance metric was used to obtain a comprehensive assessment of both accuracy and latency of the tested detectors. While the offline benchmark study identified a high-performing cluster of three detectors, online benchmarking definitively suggested the Neurokit detector as the best-performing R-peak detector, combining high inter-dataset and inter-subject performance stability with very low computational costs. Future work should extend this benchmark to embedded hardware, specifically evaluating memory footprint and power consumption constraints.

## Acknowledgment

This work was funded by the Italian Ministry of University and Research (doctoral scholarship awarded to S. Costantini).

## References

[1] J. Pan and W. J. Tompkins, “A Real-Time QRS Detection Algorithm,” *IEEE Trans. Biomed. Eng.*, 32(3):230–236, 1985.
[2] P. Hamilton, “Open source ECG analysis,” in *Comput. Cardiol.*, pp. 101–104, 2002.
[3] V. Kalidas and L. Tamil, “Real-time QRS detector using Stationary Wavelet Transform...,” in *IEEE BIBE*, pp. 457–461, 2017.
[4] M. Elgendi *et al.*, “Frequency bands effects on QRS detection,” in *Bioign. Biomed. Signal Proc.*, 2010.
[5] T. Rodrigues *et al.*, “A Low-Complexity R-peak Detection Algorithm...,” in *IEEE ICPR*, pp. 1–8, 2021.
[6] G. B. Moody and R. G. Mark, “MIT-BIH Arrhythmia Database,” PhysioNet, 1992.
[7] B. Porr and P. W. Macfarlane, “A new QRS detector stress test combining temporal jitter and F-score (JF)...,” *PLOS ONE*, 19(11):e0309739, 2024.
[8] J. P. Martínez *et al.*, “A wavelet-based ECG delineator: Evaluation on standard databases,” *IEEE Trans. Biomed. Eng.*, 51(4):570–581, 2004.
[9] I. I. Christov, “Real time electrocardiogram QRS detection using combined adaptive threshold,” *Biomed. Eng. OnLine*, 3(1):28, 2004.
[10] E. D. Lisboa, “Multi-modal behavioral biometrics based on HCI and electrophysiology,” Ph.D. diss., 2008.
[11] A. Lourenço *et al.*, “Real Time Electrocardiogram Segmentation for Finger based ECG Biometrics,” in *BIOSIGNALS*, 2012.
[12] D. Makowski *et al.*, “NeuroKit2: A Python toolbox for neurophysiological signal processing,” *Behav. Res. Methods*, 53(4):1689–1696, 2021.
[13] L. Howell and B. Porr, “High precision ECG Database with annotated R peaks,” Zenodo, 2024.
[14] N. Iyengar *et al.*, “Fantasia Database,” PhysioNet, 1994.
[15] A. I. Kalyakulina *et al.*, “LUDB: A New Open-Access Validation Tool...,” *IEEE Access*, 8:186181–186190, 2020.
[16] E. De Giovanni *et al.*, “ECG in High Intensity Exercise Dataset,” Zenodo, 2021.
[17] P. H. Charlton *et al.*, “Detecting beats in the photoplethysmogram: Benchmarking open-source algorithms,” *Physiol. Meas.*, 43(8):085007, 2022.
[18] M. Smithson and J. Verkuilen, “A better lemon squeezer? Maximum-likelihood regression...,” *Psychol. Methods*, 11(1):54–71, 2006.

 DOI 10.15203/99106-194-6-08

# Improving the stability of Venant-type approaches for the EEG and MEG forward problems by exactly matching lower-order moments

Malte B. Höltershinken, Johannes Vorwerk, Christian Engwer, Carsten H. Wolters

***Abstract*—In FEM-based approaches for the EEG and MEG forward problems, a central difficulty is simulating point dipoles in a manner that is both accurate and efficient. The so-called Venant-type approaches solve this problem by replacing the focal point dipole with a spread-out monopole cloud. This class of approaches, and especially its most recent iteration, the so-called multipole approach, has been shown to yield highly accurate simulations while simultaneously enabling rapid computations. It has, however, also been observed that, in cases where the FEM mesh in the local environment of the dipole position has a suboptimal structure, the accuracy of Venant-type approaches can occasionally degenerate. In this paper, we present a modification of the multipole approach, which, by guaranteeing that the monopole cloud and the point dipole have exactly matching monopole and dipole moments, allows the approach to still produce accurate results in cases where the local mesh structure is unfavourable.**

***Index Terms*—EEG, MEG, FEM, dipole modeling**

## I. Introduction

In EEG and MEG forward modeling, the aim is to simulate the electromagnetic signals that a given neural activity would produce. The neural activity is typically modelled as a linear combination of so-called *point dipoles*, where a point dipole consists of a position and an orientation, the latter encoding the direction and strength of current flow. Under standard assumptions, the generated electromagnetic fields depend linearly on the brain activity. The forward problems thus reduce to the task of simulating the sensor measurements corresponding to a single point dipole [1]. If one strives for a realistic head model, it becomes necessary to employ some numerical approach to perform the simulation. In this paper, we will focus on the so-called *finite element method (FEM)*, due to the ease with which it enables the inclusion of both complicated tissue structures as well as anisotropic conductivities into the model [2]. The usage of FEM approaches, however, also introduces the additional complexity of properly including the point dipole into the simulation. It has been shown that using the point dipole as-is, i.e., employing the so-called *partial integration approach*, yields comparatively inaccurate results [3]. The reason for this is that the point dipole models the current by concentrating it into a single point. If one then employs a standard FEM function space, there will be only a few test functions whose support intersects the support of the point dipole, which results in a suboptimal representation of the dipole. In Venant-type approaches, which were first introduced in [4], this problem is addressed by replacing the focal point dipole with a more distributed monopole cloud. This monopole cloud is then "seen" by a larger set of test functions, thus enabling a better representation of the dipole. If one now wants to replace the point dipole by a monopole cloud, one needs to specify the strengths of the individual monopoles in the cloud. The exact strategy for selecting these strengths is what separates the different Venant-type approaches. The most recent and accurate instance of such a strategy is the so-called *multipole approach* from [5]. There, the authors propose setting the monopole strengths in such a way that the point dipole and the monopole cloud have matching monopole, dipole, and quadrupole moments. Matching these moments translates to a system of 9 linearly independent equations, which [5] then solves by employing a regularized least squares fit. If, however, the monopole cloud does not contain enough positions to properly fit the monopole, dipole, and quadrupole moments, simultaneously trying to fit them all can reduce the accuracy of the simulation. This was, e.g., observed in [6], where the multipole approach, on the one hand, showed excellent median accuracy, while, on the other hand, yielding a noticeable amount of comparatively inaccurate simulation results for point dipoles close to tissue interfaces. The reason for this behaviour only occurring for point dipoles in close proximity to tissue interfaces is that the monopole locations contained in the interpolating monopole cloud are placed exclusively in the tissue compartment to which the original point dipole belongs, as it was shown that monopole clouds spanning multiple tissue compartments can lead to large inaccuracies in the simulation [7]. We now argue that in cases where not enough monopole positions are available to simultaneously fit the monopole, dipole, and quadrupole moments in a good way, one should prioritize fitting the monopole and dipole moments, as they predominantly influence the far-field behaviour. In this paper, we accordingly propose to modify the multipole approach in such a way that the matching of the monopole and dipole moments is explicitly enforced via linear constraints in the optimization, and the classical regularized least squares fit is only performed on the quadrupole moment, with a larger regularization parameter than in the standard multipole approach. In this way, the approach still produces accurate results in cases where only a few monopole positions are available,

M. B. Höltershinken and C.H. Wolters are with the Institute for Biomagnetism and Biosignalanalysis, University of Münster, Germany

J. Vorwerk is with the Department of Mechatronics, University of Innsbruck, Austria

C. Engwer is with the Faculty of Mathematics and Computer Science, University of Münster, Germany

 DOI 10.15203/99106-194-6-08

while simultaneously being able to benefit from matching the quadrupole moment in cases where the monopole positions allow for it. In the methods section, we will derive the formulas for the modified multipole approach. In the results section, we will then compare the modified multipole approach and the classical multipole approach in multilayer-sphere models, where analytical solutions exist, and show that the modified multipole approach increases the stability of the approach for point dipoles in close proximity to tissue interfaces.

## II. Methods

Assume we have a dipole at position $x_0 \in \mathbb{R}^3$ with moment $M \in \mathbb{R}^3$, which we want to interpolate using a monopole cloud consisting of the positions $x_1, ..., x_n \in \mathbb{R}^3$. Specifically, we have to choose monopolar loads $q_1, ..., q_n \in \mathbb{R}$ in such a way that

$$l := -\operatorname{div}(M \cdot \delta_{x_0}) \approx \sum_{i=1}^{n} q_i \cdot \delta_{x_i} =: \widehat{l},$$

where $\delta_{x_i}$ denotes the Dirac distribution at $x_i$. Once the monopolar loads have been set, Venant-type approaches then utilize the monopole cloud as the functional defining the FEM right-hand side. All instances of Venant-type approaches select their monopolar loads by specifying functions $\psi_1, ..., \psi_m$ and trying to fulfill $l(\psi_i) = \widehat{l}(\psi_i)$ for $1 \leq i \leq m$, which can equivalently be expressed as a linear equation $\langle a^{\psi_i}, q \rangle = b^{\psi_i}$ for each $i$, where $q = (q_1, ..., q_n) \in \mathbb{R}^n$. In the concrete example of the multipole approach from [5], this yields 9 linearly independent equations (1 for the monopole moment, 3 for the dipole moment, and 5 for the symmetric traceless quadrupole tensor). After gathering the linear equations into a matrix $X^\intercal = (a^{\psi_1}, ..., a^{\psi_m}) \in \mathbb{R}^{n \times m}$ and a right-hand side $t = (b^{\psi_1}, ..., b^{\psi_m}) \in \mathbb{R}^m$, and defining a weight matrix $\overline{W} \in \mathbb{R}^{n \times n}$, the multipole weights in [5] are then defined by

$$q^{\text{multipole}} := \operatorname*{arg\,min}_{q \in \mathbb{R}^n} \|\overline{X} \cdot q - \overline{t}\|_2^2 + \lambda \cdot \|\overline{W} \cdot q\|_2^2,$$

where $\overline{X}$ and $\overline{t}$ arise from $X$ and $t$ by scaling each row with a factor depending on the corresponding $\psi_i$, and $\lambda$ is a regularization parameter.

We can now assume that the first four rows of $\overline{X}$ correspond to the monopole and dipole moment conditions. Then, let $\overline{X}^\intercal = (A, B)$, with $A \in \mathbb{R}^{n \times 4}$ and $B \in \mathbb{R}^{n \times (m-4)}$. Define $\overline{t} = (a, b)$ accordingly. As motivated in the introduction, we now want to exactly fulfill the conditions expressed by $A$ and $a$, and only perform a regularized least squares fit over the conditions expressed by $B$ and $b$. Explicitly, we define the modified multipole weights by

$$q^{\text{mod. multipole}} := \operatorname*{arg\,min}_{A^\intercal \cdot q = a} \|B \cdot q - b\|_2^2 + \lambda \cdot \|\overline{W} \cdot q\|_2^2.$$

For the multipole approach, we have $b = 0$, as shown in [5]. A standard computation using Lagrange multipliers, as, e.g., described in [8], then yields that, if we define

$$N = B \cdot B^\intercal + \lambda \cdot \overline{W}^\intercal \cdot \overline{W} \in \mathbb{R}^{n \times n},$$

we have

$$q^{\text{mod. multipole}} = N^{-1} \cdot A \cdot \left(A^\intercal \cdot N^{-1} \cdot A\right)^{-1} \cdot a.$$

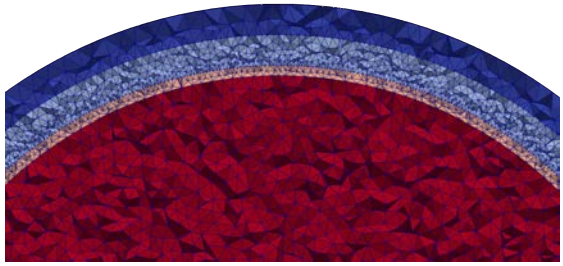

Fig. 1: *mesh_init*, consisting of ca. 800k nodes and ca. 5 million elements.

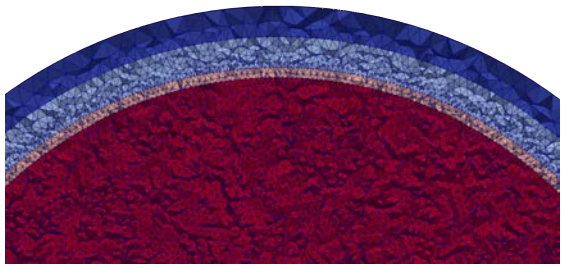

Fig. 2: *mesh_brain*, consisting of ca. 1.8 million nodes and ca. 11 million elements. *mesh_brain* was created from *mesh_init* by refining the innermost compartment.

The modified multipole approach was implemented into the DUNEuro toolbox [2], which we used for our numerical experiments. In these experiments, we compared the accuracy of the multipole approach and the modified multipole approach in a multilayer-sphere model with standard parameters, as, e.g., given in [5]. For this geometry, we then employed two different FEM meshes from [9]. In detail, we used one mesh with a comparatively low brain resolution, called *mesh_init*, and one mesh with a high brain resolution, called *mesh_brain*. These meshes are visualized in figures 1 and 2. For a given *eccentricity*[1], we then generated 1000 random radially oriented dipoles at that eccentricity, and, for each dipole, computed the multipole EEG forward solutions, the modified multipole EEG forward solutions, and the relative errors with respect to the analytical solutions [10] in both meshes.

## III. Results

In a first step, we compared the multipole approach and the modified multipole approach using the standard parameters as suggested by [5], i.e., we used a regularization parameter of $\lambda = 10^{-6}$. The results for *mesh_init* can be seen in figure 3. There, we see that both the multipole approach and the modified multipole approach yield a large number of outliers at high eccentricities.

The reason for this is that setting $\lambda = 10^{-6}$ is too low for the regularization to have a beneficial effect. If we instead

[1]The eccentricity of a point in the multilayer-sphere mesh is given by the quotient of its distance to the sphere center and the radius of the brain, i.e., the innermost compartment.

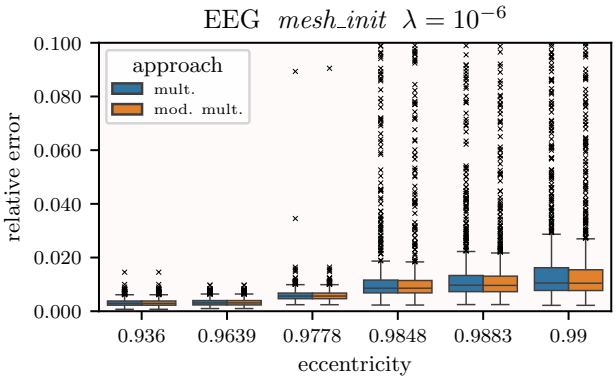


Fig. 3: Comparison of the multipole approach ("mult.") and the modified multipole approach ("mod. mult.") using a low regularization parameter.

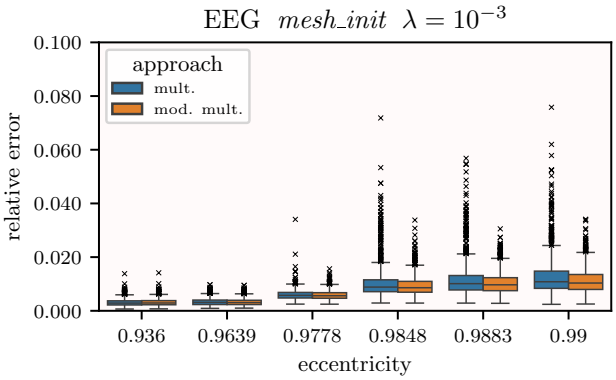


Fig. 4: Comparison of the multipole approach ("mult.") and the modified multipole approach ("mod. mult.") using a high regularization parameter.

choose $\lambda = 10^{-3}$, as shown in figure 4, the amount of outliers for both the multipole approach as well as for the modified multipole approach is strongly decreased. Furthermore, we see that the modified multipole approach has a small but noticeable edge over the multipole approach at high eccentricities.

Finally, if we compare the accuracy of the multipole approach using $\lambda = 10^{-6}$ with the modified multipole approach using $\lambda = 10^{-3}$ in *mesh_brain*, where there are always enough monopole positions available to easily fit the monopole moment, dipole moment, and quadrupole moment, we see that both methods show an almost identical performance.

We also ran these comparisons for tangential dipoles and for the MEG forward problem, which yielded the same qualitative result.

## IV. Conclusion

The numerical experiments showed that the standard choice for the Venant regularization parameter, i.e., $\lambda = 10^{-6}$, is, in general, too low to have a beneficial effect in situations where the interpolating monopole cloud contains only a few positions, and should be increased to $\lambda = 10^{-3}$. We saw that, for this larger regularization parameter, the modified multipole approach shows better numerical stability for sources close to tissue interfaces, and otherwise matches the performance of the standard multipole approach.

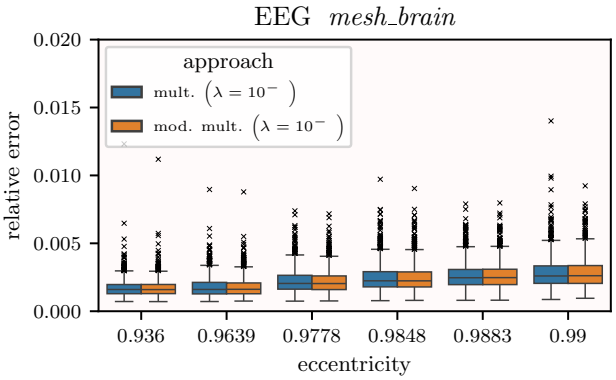


Fig. 5: Comparison of the multipole approach ("mult.") using a low regularization and the modified multipole approach ("mod. mult.") using a high regularization. Note the different scale used for the errors than in the other figures.

## Acknowledgment

This work was supported by the Deutsche Forschungsgemeinschaft (DFG), projects WO1425/10-2 and WO1425/11-1 (MBH, CHW), by ERA PerMed as project ERAPERMED2020-227, PerEpi (Bundesministerium für Gesundheit (BMG), project ZMI1-2521FSB006 (MBH, CHW)), by the DAAD, project 57756407 (MBH, CHW), by the DFG through the Cluster of Excellence "Mathematics Münster: Dynamics - Geometry - Structure" (EXC 2044/2–390685587) (MBH, CE), and by the Austrian Science Fund (FWF), Austria [ 10.55776/P35949] (JV).

## References

[1] M. Hämäläinen *et al.*, "Magnetoencephalography—theory, instrumentation, and applications to noninvasive studies of the working human brain," *Rev. Mod. Phys.*, vol. 65, pp. 413–497, 2 1993. DOI: 10.1103/RevModPhys.65.413.

[2] S. Schrader *et al.*, "DUNEuro—A software toolbox for forward modeling in bioelectromagnetism," *PLOS ONE*, vol. 16, no. 6, pp. 1–21, 2021. DOI: 10.1371/journal.pone.0252431.

[3] S. Pursiainen *et al.*, "Electroencephalography (EEG) Forward Modeling via H(div) Finite Element Sources with Focal Interpolation," *Physics in medicine and biology*, vol. 61, 2016. DOI: 10.1088/0031-9155/61/24/8502.

[4] H. Buchner *et al.*, "Inverse localization of electric dipole current sources in finite element models of the human head," *Electroencephalography and Clinical Neurophysiology*, vol. 102, no. 4, pp. 267–278, 1997. DOI: https://doi.org/10.1016/S0013-4694(96)95698-9.

[5] J. Vorwerk *et al.*, "The multipole approach for EEG forward modeling using the finite element method," *NeuroImage*, vol. 201, p. 116 039, 2019. DOI: https://doi.org/10.1016/j.neuroimage.2019.116039.

[6] M. B. Höltershinken *et al.*, "The Local Subtraction Approach for EEG and MEG Forward Modeling," *SIAM Journal on Scientific Computing*, vol. 47, no. 1, B160–B189, 2025. DOI: 10.1137/23M1582874. eprint: https://doi.org/10.1137/23M1582874.

[7] T. Medani *et al.*, "FEM Method for the EEG Forward Problem and Improvement Based on Modification of the Saint Venant's Method," *Progress in Electromagnetics Research*, vol. 153, pp. 11–22, 2015. DOI: 10.2528/PIER15050102.

[8] S. Boyd *et al.*, *Convex optimization*. 2004.

[9] M. B. Höltershinken *et al.*, *Data used for validation of the Local Subtraction Approach*, 2024. DOI: 10.5281/zenodo.12575552.

[10] J. de Munck *et al.*, "A fast method to compute the potential in the multisphere model (EEG application)," *IEEE Transactions on Biomedical Engineering*, vol. 40, no. 11, pp. 1166–1174, 1993. DOI: 10.1109/10.245635.

 DOI 10.15203/99106-194-6-09

# A Polymer Optical Fiber-Based Approach for Capturing LED-Specific Photoplethysmography Channels from Wearables

L. Keck, R. Drath, B. Dömer

***Abstract*—In this contribution, the authors present a novel interface concept for the validation of wearable health trackers. Smartwatches and fitness trackers commonly measure pulse rate and oxygen saturation by emitting green, red, and infrared light. Validating the accuracy of these sensors is challenging and typically requires extensive testing, either on human subjects or via simulators that both read and stimulate the device. A key difficulty is connecting a simulator to the wearable, as LED–photodiode geometries vary between manufacturers. To address this, we developed a generic optical interface that decouples the simulator from the wearable's geometry while still enabling measurement of wavelength-resolved illumination patterns and LED-specific intensities. The proposed setup uses polymer optical fibers, optical splitters, and wavelength-selective filters to guide and separate light from individual LEDs. The paper presents the design, implementation, and validation of this approach, demonstrating its suitability for systematic testing of wearable PPG sensors.**

***Index Terms*—Polymer Optical Fiber, Photoplethysmography, LED drive patterns, Simulator, Wearables, Smartwatch**

## I. Introduction and Requirements

Health tracking has become a global trend, and wearable devices now monitor pulse and other physiological parameters continuously. Wearables typically rely on photoplethysmography (PPG), where photodiodes (PD) detect changes in light intensity after the skin is illuminated by LEDs at specific wavelengths [1]. Most devices use green, red, and infrared LEDs, but they differ in the number and type of LEDs and PDs, their spatial arrangement, and LED drive patterns [2]. These design choices affect the detected signal and reflect trade-offs between signal-to-noise ratio (SNR) and power consumption. To validate the accuracy of these wearables without extensive human testing, optical simulators are used that both read the LED drive patterns and provide controlled optical feedback to the PDs.

PPG sensors operate in transmission mode, with LED and PD on opposite sides of the tissue, or in reflectance mode, where both lie on the same surface. In reflectance mode, light paths depend strongly on the exact LED–PD geometry, making it essential to know which LED is active at any moment. Individual emitters and PDs, however, cannot be easily isolated. This variability in geometry and illumination patterns poses significant challenges for simulator design, since accurate simulation of tissue reflectance requires knowledge of the wavelength-specific emission of each LED over time. A flexible interface is therefore required that captures wavelength-specific light output from each LED and adapts to diverse wearable configurations while minimizing cost.

The question is: How can a simulator interface be designed to capture wavelength-specific light output from each LED while remaining compatible with wearables of diverse geometries and LED arrangements?

To address this challenge, the interface must (a) be able to distinguish the output of individual LEDs, (b) even when multiple wavelengths are emitted from a single chip, and (c) accurately separate and measure different wavelengths. It should also (d) adapt to varying LED–photodiode layouts without requiring modifications to the wearable device, while keeping adaptation effort and cost low.

L. Keck, R. Drath and B. Dömer are with the Department of IoS$^3$, Pforzheim University, Germany, e-mail: Luca.Keck@hs-pforzheim.de.

## II. Related Work

There are several methods for simulation approaches for photoplethysmography sensors on wearable devices. One example is the optical signal simulator developed by Pittella et al. [3], which uses a single optical channel that merges all emitted wavelengths into a single return path. The system simulates green, red, and infrared illumination and feeds the aggregated optical signal back to the device, which is positioned at a fixed distance [3].

The commercial reflectance $SpO_2$ module PPG-2RS-940 (WhaleTeq, Taiwan) uses a dedicated photodetector region to measure the intensity of the combined output from all LEDs of the device under test (DUT). The measured signal is then used to generate a simulated optical output that is emitted uniformly toward all PDs through a single illumination area [4]. Mechanical adapters are used to isolate LEDs and PDs by blocking optical crosstalk and directing the light toward designated regions, enabling compatibility with multiple sensor geometries. However, the system does not provide LED-specific mapping or photodiode-specific illumination.

In summary, both approaches respond only to the total optical emission of the DUT and return a single feedback signal, without distinguishing individual LEDs or generating photodiode-specific optical feedback, which does not fulfill requirement (a).

## III. Idea and Concept

The authors propose decoupling the wearable device from the simulator by introducing individual optical channels implemented using a polymer optical fiber (POF) with a $1000\,\mu m$

 DOI 10.15203/99106-194-6-09

core (GH4001, Industrial Fiber Optics, USA). The polished fiber tip is positioned as close as possible to the LED to maximize optical coupling while avoiding any modification to the wearable device.

The POF exhibits wavelength-dependent attenuation of approximately $0.1\,\mathrm{dB/m}$ at 540 nm (green), $0.2\,\mathrm{dB/m}$ at 660 nm (red), and $4\,\mathrm{dB/m}$ at 940 nm (infrared),which must be taken into account when selecting the fiber length [5].

Because flexible POFs, glued into designated holes, can route emitted light independently of the wearable's geometry, the illumination from each LED can be guided to the simulator regardless of its physical location on the device. For each wearable, a custom fiber holder can be fabricated to ensure proper alignment, block ambient light, and position the fiber optimally for collecting LED emission and delivering simulated light to the device's photodiode, as illustrated in Figure 1.

This approach enables low-cost and low-effort adaptation to different LED–PD layouts without requiring any modification to the wearable device itself, which fulfills requirement (d).

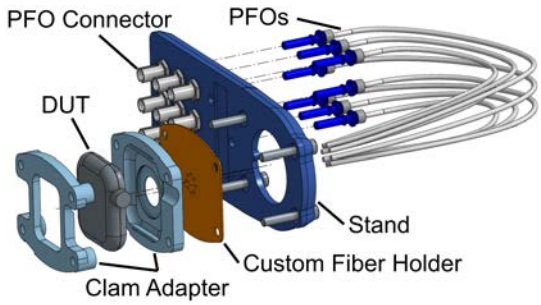


Fig. 1: Decoupling setup using LED-PD-specific POFs

### A. Wavelength discrimination

Many PPG sensors employ multi-wavelength LEDs that emit multiple wavelengths from a single chip. Because these wavelengths originate from nearly the same spatial emission region, they cannot be physically separated using individual optical fibers. To enable wavelength discrimination as required in requirement (c), all emitted wavelengths are initially coupled into a single POF core.

To recover the wavelength-specific illumination intensities, the combined optical signal must be spectrally separated before detection. This is necessary because the simulator uses PDs that measure only total optical power, providing no information about the wavelength of the incoming light.

To separate the wavelengths, an n:3 optical splitter is used to generate three independent channels. This configuration allows recording signals from any number of LEDs required for the measurement. When combined with wavelength-selective bandpass filters placed in front of each PD, the setup achieves effective spectral separation, as illustrated in Figure 2.

For the infrared channel, an acrylic filter that blocks visible light up to 700 nm is positioned directly in front of the PD. For the green and red channels, any residual infrared light is attenuated by adding an extra segment of POF. As noted earlier, the POF exhibits substantially higher transmission loss at infrared wavelengths compared to visible wavelengths. Consequently, the additional fiber length reduces the infrared component to negligible levels. To filter the red and green channels, colored optical foils corresponding to each wavelength are stacked between the POF and the PD to achieve clear differentiation.

The complete measurement and filtering setup for a wearable device containing four LEDs is shown in Figure 2.

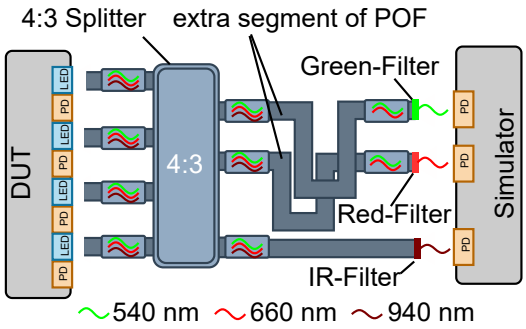


Fig. 2: Measurement setup illustrating wavelength separation of four LEDs using a 4:3 optical splitter and wavelength-selective filters.

The inclusion of additional optical connectors, filters, and the splitter introduces significant attenuation, resulting in substantially lower readings on the PD, as shown in the wavelength-dependent LED pattern in Figure 3.

The total attenuation of each optical path from the wearable's LED to the simulator's PD, $A_{\mathrm{tot}}$, is wavelength-dependent and can be expressed as:

$$A_{\mathrm{tot}} = A_{\mathrm{cpl}} + A_{\mathrm{con}}(n) + A_{\mathrm{POF}}(l,\lambda) + A_{\mathrm{split}} + A_{\mathrm{filt}}(\lambda) \quad (1)$$

where $A_{\mathrm{cpl}}$ is the coupling loss from the LED into the fiber core; $A_{\mathrm{con}}(n)$ is the total connection loss introduced by $n$ POF connectors, including those used to interface the splitter and the detachable fiber holder. These connectors enable swapping between different wearables without modifying the underlying measurement setup; $A_{\mathrm{POF}}(l,\lambda)$ is the fiber attenuation over length $l$; $A_{\mathrm{split}}$ represents the splitter loss, accounting for all optical power not directed into the selected output channel; and $A_{\mathrm{filt}}(\lambda)$ is the attenuation introduced by the wavelength-selective filter.

## IV. Measurements

The proposed approach for capturing LED-specific photoplethysmography channels from wearables was validated using an Apple Watch Series 10.

The three available channels, without optical filtering, were connected to the top, right, and bottom LEDs of the Apple Watch (as viewed from the front), revealing the position-dependent LED pattern shown in Figure 3. The plot shows distinct groups of LED bursts (highlighted in red), each consisting of pulses approximately $110\,\mu\mathrm{s}$ in duration, separated by gaps of similar length. Between groups, a pause of about $3\,\mathrm{ms}$ occurs before the next series begins.

 DOI 10.15203/99106-194-6-09

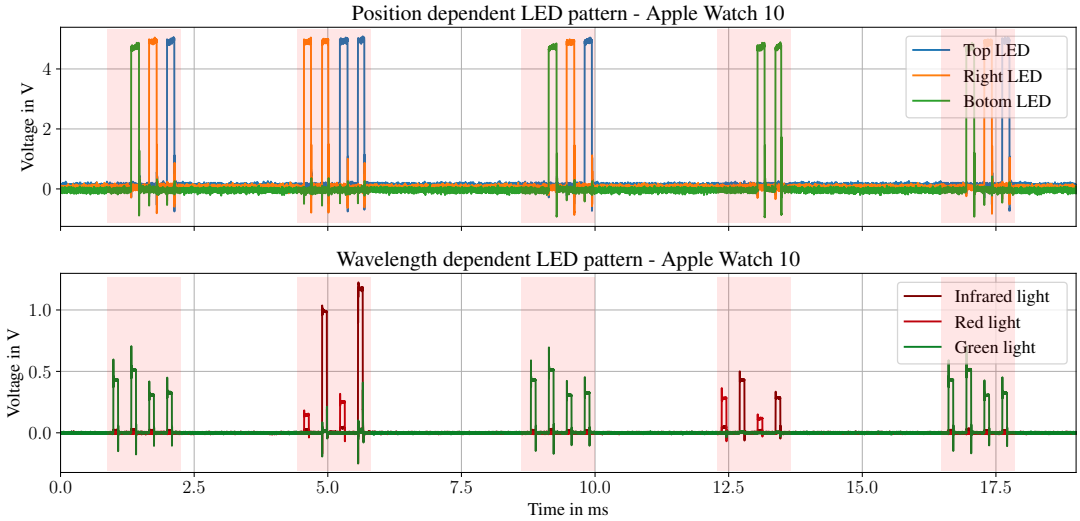


Fig. 3: LED driver pattern of the Apple Watch 10: top, right, and bottom LEDs connected directly to the simulator without optical filters (top), and all four LEDs from the Apple Watch filtered by wavelength (bottom).

Spatially resolved measurements across the three channels highlight differences between groups: in groups 1, 3, and 5, each monitored LED pulses once, whereas in group 2, two opposite LEDs pulse twice consecutively. Groups 1, 3, and 5 correspond to a counterclockwise sequence of single pulses, whereas groups 2 and 4 contain double pulses from opposite LED pairs (right–top or bottom–left).

To analyze wavelength content, the four LEDs were connected to a 4:3 optical splitter and measured using infrared and visible light filters. With this configuration, each PD received light from all four LEDs filtered by wavelength, enabling a wavelength-specific analysis that can be matched to the first, spatially resolved measurement. The resulting data are shown in the lower half of Figure 3. The wavelength-dependent signals indicate that groups 1, 3, and 5 contain green light, whereas the pulses in groups 2 and 4 contain alternating red and infrared light. The differences in intensity levels arise from the splitter's imbalance and from variations in coupling and connection losses.

## V. Discussion

This system enables recording the LED-specific pulse behavior of wearable PPG sensors, revealing a structured LED-driving sequence, addressing Requirement (a). Extending the system beyond three channels would allow characterization of wearables with more than three LEDs and simultaneous discrimination of multiple wavelengths. A remaining challenge is the implementation of the simulator's optical output. This requires generating physiologically meaningful feedback signals for each photodiode based on the measured input channels. Achieving this capability would allow investigation of how the device responds to varying SNR conditions and adapts its LED-driving pattern. Further optimization of channel loss equalization is necessary to accurately quantify the relative light intensities at the photodiodes.

## VI. Conclusion

The presented approach demonstrates that LED-driving signals from a wearable PPG sensor can be captured via POFs and prefiltered using wavelength-selective optical filters, enabling separation of individual wavelength components, as required in (a). This method further allows characterization of the behavior of each LED and its wavelength-dependent output, directly addressing (c), while operating without any modifications to the wearable device itself in accordance with (d).

The flexible setup, which can be adapted to a variety of LED–photodiode configurations in line with (d), provides a foundation for developing more physiologically accurate optical simulators. Future enhancements will support more precise investigations of sensor performance, LED-driving strategies, and signal-to-noise behavior under realistic conditions, even when multiple wavelengths originate from a single LED chip, as captured in (b).

## References

[1] P. A. Kyriacou und J. Allen (Hrsg.), *Photoplethysmography: Technology, Signal Analysis and Applications*. Elsevier Academic Press, London, 2022. ISBN: 978-0-12-823374-0, 978-0-12-823525-6.

[2] K. B. Kim and H. J. Baek, Photoplethysmography in Wearable Devices: A Comprehensive Review of Technological Advances, Current Challenges, and Future Directions, *Electronics*, vol. 12, no. 13, p. 2923, 2023, doi: 10.3390/electronics12132923.

[3] E. Pittella and O. Testa *et al.*, An Optical Signal Simulator for the Characterization of Photoplethysmographic Devices, *Sensors*, vol. 24, no. 3, p. 1008, 2024. doi: 10.3390/s24031008.

[4] WhaleTeq Co., Ltd., Oximeter Device Testing, available online: https://www.whaleteq.com/en/product/15-oximeter-device-testing/list, accessed Nov. 14, 2025.

[5] i–FiberOptics, *Attenuation of Optical Fiber Communication*. Available online: https://i-fiberoptics.com/pdf/attenuation.pdf. Accessed: Nov. 14, 2025.

 DOI 10.15203/99106-194-6-10

# Localization of a Magnetic Nanoparticle Bolus in Motion Using Magnetorelaxometry Imaging

S. Arsalani, R. Körber, D. Baumgarten and F. Wiekhorst

***Abstract*—Magnetorelaxometry imaging (MRXI) is a non-invasive technique used for localization and quantification of magnetic nanoparticles (MNPs) within a defined region. In this work, we investigated the localization of a fluid MNP bolus in motion with an iron concentration of 0.70 mg/ml using an MRXI system, which was designed and developed in our previous work to track the MNP movement. In this work, the experiments were performed at different flow velocities ranging from 2.9 mm/s to 10.6 mm/s, within a tubular phantom. The reconstructed images of the MNP bolus show that the localization aligns well with the ground truth images at all flow velocities with a minimum temporal resolution of 1.1 s and an image resolution of approximately (6.3 × 6.3 × 2.1) mm$^3$. This approach has potential for the diagnosis of life-threatening conditions such as cerebral aneurysms using a low iron concentration.**

***Index Terms*—magnetic nanoparticles, magnetorelaxometry imaging, flow, cerebral aneurysms**

## I. Introduction

Cerebral aneurysms are a serious condition, which develop as abnormal balloon-like bulges in cerebral arteries, and their rupture can lead to some life-threatening complications. Several clinical imaging techniques are used for the diagnosis of cerebral aneurysms such as digital subtraction angiography (DSA), computed tomography angiography (CTA) and magnetic resonance angiography (MRA) [1, 2, 3]. Although each of these clinical techniques offers specific advantages, they also have some drawbacks. For example, DSA offers high spatial and temporal resolution but is invasive. CTA is a rapid and non-invasive technique but involves ionizing radiation. MRA is non-invasive and does not use ionizing radiation but often requires the use of gadolinium contrast agents, which are not suitable for kidney patients.

Recently, a preclinical magnetic particle imaging (MPI) scanner, with high spatial and temporal resolution, successfully visualized flow of magnetic nanoparticles (MNPs) in a 3D-printed, patient-specific cerebral aneurysm phantom [4]. However, when applied to human sized head phantoms, this system provides currently only 2D images [5], which can restrict its diagnostic efficacy for cerebral aneurysms.

Another potential technique for the diagnosis of cerebral aneurysms is magnetorelaxometry imaging (MRXI), a novel and non-invasive modality, which is commonly used to localize and quantify the spatial distribution of MNPs within a specific object [6].

Very recently, we demonstrated the capability of MRXI for monitoring the flow of an MNP bolus using a modular 72-channel superconducting quantum interference device (SQUID) sensor system [7]. These experiments were conducted with different iron concentrations $c$(Fe) ranging from 1.40 mg/ml to 5.58 mg/ml and flow velocities from 2.9 mm/s to 8.2 mm/s. In the present work, we further reduced $c$(Fe) to approximately 0.70 mg/ml and increased the flow velocity up to 10.6 mm/s in order to assess the ability of the system to monitor the MNP bolus with lower $c$(Fe) and higher velocity, conditions that are crucial for the diagnosis of cerebral aneurysms.

## II. Methods

### A. MRXI Setup

Our MRXI setup consists of a 72-channel low-$T_c$ current sensing SQUID system housed in a liquid helium Dewar, a tubular phantom with 2.1 mm diameter, and six spiral excitation coils, each with 36 turns and an outer diameter of 12 mm. The generated magnetic field was about 2.28 mT at the center of the excitation coil, using a current of 0.6 A. The tubular phantom was fixed on top of the excitation coils and positioned below the Dewar with vertical distance of approximately 45 mm. An overview of the setup is depicted in Fig. 1.

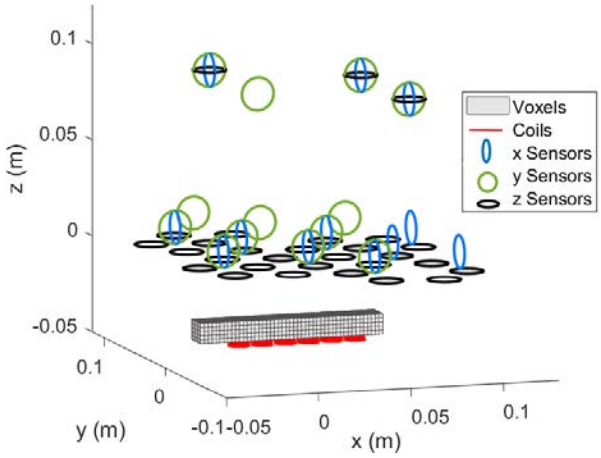


Fig. 1. An overview of the developed MRXI setup. The schematic illustrates the voxels with size of (2.1 × 2.1 × 2.1) mm$^3$ in gray and the excitation coils in red. The sensor positions for $x$-, $y$-, and $z$-directions are shown in blue, green and black, respectively.

S. Arsalani, R. Körber and F. Wiekhorst are with Physikalisch-Technische Bundesanstalt, Abbestrasse 2-12, D-10587 Berlin, Germany, e-mail: soudabeh.arsalani@ptb.de.

D. Baumgarten is with Biomedical Engineering Group, Department of Mechatronics, University of Innsbruck, Innsbruck, Austria.

 DOI 10.15203/99106-194-6-10

### B. *Measurement procedure*

In these experiments, about 90 µl of a fluid MNP suspension in water (IONP, Micromod Partikeltechnologie GmbH) with a hydrodynamic diameter of 250 nm and $c$(Fe) of 0.70 mg/ml was injected into the tubular phantom then pushed through the tube at four different flow velocities: 2.9, 4.8, 8.2 and 10.6 mm/s. The MNP bolus in motion was sequentially magnetized using the spiral coils for a magnetization time $t_{mag}$ = 0.5 s. Then, after switching off the magnetization fields and a short delay time of 100 µs, the relaxation signals of the MNP bolus were recorded by the SQUID sensors for a specific measurement time $t_{meas}$, see Table I. This was followed by a pause time $t_{pause}$ = 0.1 s. For each experiment, the number of shots applied to each coil was selected based on the flow velocity to ensure that the MNP bolus was mostly above the activated coil as it passed through the tube phantom.

TABLE I: MAGNETIZATION TIME, MEASUREMENT TIME AND THE NUMBER OF SHOTS FOR DIFFERENT FLOW VELOCITIES.

| Flow velocities (mm/s) | $t_{mag}$ (s) | $t_{meas}$ (s) | Number of shots |
|---|---|---|---|
| 2.9 | 0.5 | 0.8 | 3 |
| 4.8 | 0.5 | 0.8 | 2 |
| 8.2 | 0.5 | 0.8 | 1 |
| 10.6 | 0.5 | 0.5 | 1 |

### C. *Image reconstruction*

First, we define the region of interest (ROI), which considered approximately (96.6 × 14.7 × 10.5) mm$^3$ and is divided into $N_v$ voxels with a voxel size of (2.1 × 2.1 × 2.1) mm$^3$. The magnetic field generated at the center of each voxel $v$ is then calculated based on the excitation current and the geometry of the excitation coil. Each signal source, which is magnetized by the magnetic field, is modeled as a dipole at its voxel center $r_v$. The magnetic flux density $B_s$, detected by a sensor $s$ at position $r_s$, can be calculated as described in [8]:

$$B_s(t) = \frac{\mu_0}{4\pi} \sum_{v=1}^{N_v} \left[ \frac{3(\mathbf{n}_s^T \mathbf{r}_{s,v}) \mathbf{r}_{s,v}^T}{\| \mathbf{r}_{s,v} \|^5} - \frac{\mathbf{n}_s^T}{\| \mathbf{r}_{s,v} \|^3} \right] \mathbf{H}_v \chi \xi(t) x_v \quad (1)$$

where $\mu_0$ denotes the magnetic permeability of free space, $\mathbf{n}_s$ is the sensor orientation, and $\mathbf{r}_{s,v}$ the vector pointing form voxel $v$ to sensor $s$. $\mathbf{H}_v$ is the magnetic field generated by the excitation coil in voxel $v$, calculated by Biot–Savart's law. $\chi$ is the magnetic susceptibility, $\xi(t)$ is the relaxation function, and $x_v$ is the concentration of MNPs in the individual voxel $v$. The relaxation amplitudes, which are recorded by all $N_s$ sensors throughout the $N_a$ activation sequences, can be written as:

$$\mathbf{\Delta B} = \mathbf{L}\mathbf{x} \quad (2)$$

where $\Delta\mathbf{B} \in R^{N_s N_a}$ is a vector of relaxation amplitudes, $\mathbf{x} \in R^{N_v}$ is a vector of MNP concentration of all voxels and $\mathbf{L} \in R^{N_s N_a \times N_v}$ contains all model parameters.

By solving the ill-conditioned inverse problem of the forward model, eq. (2), the spatial MNP distributions can be reconstructed. Here, we use Tikhonov regularization with added non-negativity constraint to solve the inverse problem:

$$\tilde{\mathbf{x}} = \text{argmin} \|\mathbf{L}\mathbf{x} - \mathbf{\Delta B}\|_2^2 + \alpha \|\mathbf{\Gamma}\mathbf{x}\|_2^2 \quad (3)$$

where $\alpha \in R$ is the regularization parameter and $\mathbf{\Gamma}$ a weighting matrix comprised of the identity matrix. The reconstructions were performed using a single activation (shot) of each excitation coil and the regularization parameters were automatically selected based on the corner of the $L$-curve method using Tikhonov regularization [9, 7].

Here, Pearson correlation coefficient ($CC$) was used to evaluate the qualitative similarity between the reconstructed and the ground truth MNP distribution, where a $CC$ value of 0 indicates no correlation and $CC$ value of 100 % indicates a perfect correlation between two images [10].

## III. RESULTS

The reconstructed and ground truth images of the MNP bolus with a $c$(Fe) of 0.70 mg/ml and the maximum flow velocity of 10.6 mm/s within the tubular phantom, are shown in Fig. 2 (a), for coil 2-shot 1 with a $CC$ value of 36.4 %. The image was reconstructed using regularization parameter selected based on the L-curve corner. The selected point is marked red in Fig. 2 (b).

From the reconstructed images at all velocities, we infer a spatial resolution of approximately (6.3 × 6.3 × 2.1) mm$^3$. All reconstructed images show that the position of the MNP bolus aligned well with the ground truth images in different positions within tube phantom. Nevertheless, the reconstructed image appears with a shorter length than the ground truth. This shortening can be attributed to the use of compact excitation coils, which generate a highly localized magnetic field, resulting in magnetizing only a fraction of the bolus located above the coil.

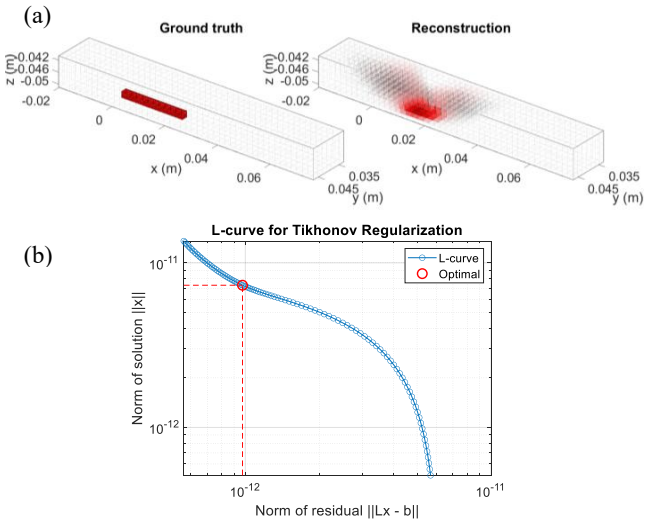


Fig. 2. (a) The reconstructed and ground truth images of the MNP bolus with $c$(Fe) of 0.70 mg/ml at a flow velocity of 10.6 mm/s, within the tubular phantom, for the activation of coil 2-shot 1. (b) The calculated L-curve for a single shot with selected regularization parameter in red.

The maximum *CC* ($CC_{max}$) for all flow velocities, used as the figure of merit, allows comparison between the different velocities and is shown in Table II. The MNP bolus with the flow velocity of 4.8 mm/s exhibits the highest $CC_{max}$, which can be attributed to a more precise synchronization between the motion of the MNP bolus and the MRXI sequence.

TABLE II: THE MAXIMUM CORRELATION COEFFICIENT, $CC_{MAX}$, AND TEMPORAL RESOLUTION FOR DIFFERENT FLOW VELOCITIES OF THE MNP BOLUS.

| Flow velocities (mm/s) | $CC_{max}$ (%) | Temporal resolution (s) |
|---|---|---|
| 2.9 | 38.0 | 1.4 |
| 4.8 | 42.0 | 1.4 |
| 8.2 | 37.6 | 1.4 |
| 10.6 | 36.5 | 1.1 |

The results in Table II show a slightly lower $CC_{max}$ for the MNP bolus with $c$(Fe) = 0.70 mg/ml compared to higher $c$(Fe) (1.4 mg/ml to 5.58 mg/ml) reported in our previous work [6]. This reduction can be attributed to a reduced signal-to-noise ratio which is due to the smaller relaxation amplitude resulting from the lower concentration of the bolus. Nevertheless, in this study, we improved the temporal resolution for the highest velocity (10.6 mm/s) to 1.1 s (including $t_{mag}$ = 0.5 s, $t_{meas}$ = 0.5 s, and $t_{paus}$ = 0.1 s) by reducing $t_{meas}$, which is very beneficial for the potential diagnosis of cerebral aneurysms.

## IV. CONCLUSION

In this work, we assessed the ability of the MRXI system, which was previously designed and developed, to track the movement of a MNP bolus under a low $c$(Fe). The experiments were performed using a fluid MNP bolus with an iron concentration of 0.70 mg/ml at flow velocities ranging from 2.9 mm/s to 10.6 mm/s. The reconstructed results at various locations within a tube phantom showed that this MRXI system is capable of tracking and reconstructing a moving MNP bolus with the iron concentration of 0.70 mg/ml at all flow velocities. The system achieved an image resolution of approximately (6.3 × 6.3 × 2.1) $mm^3$ and a minimum temporal resolution of 1.1 s.

## ACKNOWLEDGMENT

Financial support by the German Science Foundation (DFG), Project ‘quantMRX’, Grant 428329263, WI4230/4-1, by PTB, Project ‘Next generation PTB SQUIDs’ and by Austrian Science Fund (FWF), Grant No. I 4357-B, is gratefully acknowledged.

## REFERENCES

[1] X. Wang, J. C. Benson, B. Jagadeesan, and A. McKinney, “Giant cerebral aneurysms: Comparing CTA, MRA, and digital subtraction angiography assessments,” *Journal of Neuroimaging*, vol. 30, pp. 335–341, 2020. doi: 10.1111/jon.12712.

[2] C. Kouskouras, A. Charitanti, C. Giavroglou, N. Foroglou, P. Selviaridis, V. Kontopoulos, and A. S. Dimitriadis, “Intracranial aneurysms: evaluation using CTA and MRA. correlation with DSA and intraoperative findings,” *Neuroradiology*, vol. 46, pp. 842–850, Sep. 2004. doi: 10.1007/s00234-004-1259-2.

[3] E. Z. Kapsalaki, C. D. Rountas, and K. N. Fountas, “The role of 3 tesla MRA in the detection of intracranial aneurysms,” *International Journal of Vascular Medicine*, vol. 2012, pp. 1–9, 2012. doi: 10.1155/2012/792834.

[4] O. Kosch, A. F. Sanches, Y. Özpeynirci, T. Liebig, and F. Wiekhorst, “Milisecond magnetic particle imaging of a 3d printed patient-specific cerebral aneurysm phantom with flow diverter stent,” *Transactions on Additive Manufacturing Meets Medicine*, p. Vol. 5, 2023. doi: 10.18416/AMMM.2023.2309808.

[5] E. Mattingly, M. Åšliwiak, E. Mason, J. Chacon-Caldera, A. Barksdale, F. H. Niebel, K. Herb, M. Graeser, and L. L. Wald, “Design, construction and validation of a magnetic particle imaging (MPI) system for human brain imaging,” *Physics in Medicine & Biology*, vol. 70, p. 015019, 2025. doi: 10.1088/1361-6560/ad9db0.

[6] M. Liebl, U. Steinhoff, F. Wiekhorst, J. Haueisen, and L. Trahms, “Quantitative imaging of magnetic nanoparticles by magnetorelaxometry with multiple excitation coils,” *Physics in Medicine & Biology*, vol. 59, pp. 6607–6620, 2014. doi: 10.1088/0031-9155/59/21/6607.

[7] S. Arsalani, F. Wiekhorst, A. Jaufenthaler, D. Baumgarten, and R. Körber, “Monitoring the movement of a bolus of magnetic nanoparticles by magnetorelaxometry imaging,” *Physics in Medicine & Biology*, vol. 70, p. 165003, 2025. doi: 10.1088/1361-6560/adf40f.

[8] S. Arsalani, P. Radon, P. Schier, A. Jaufenthaler, M. Liebl, D. Baumgarten, and F. Wiekhorst, “Developing magnetorelaxometry imaging for human applications,” *Physics in Medicine & Biology*, vol. 67, p. 225007, 2022. doi: 10.1088/1361-6560/ac9c41.

[9] A. Jaufenthaler, T. Sander, P. Schier, K. Pansegrau, F. Wiekhorst, and D. Baumgarten, “Human head sized magnetorelaxometry imaging of magnetic nanoparticles with optically pumped magnetometers – A feasibility study,” *Journal of Magnetism and Magnetic Materials*, vol. 596, p. 171983, 2024. doi: 10.1016/j.jmmm.2024.171983.

[10] P. Schier, A. Jaufenthaler, M. Liebl, S. Arsalani, F. Wiekhorst, and D. Baumgarten, “Human-sized quantitative imaging of magnetic nanoparticles with nonlinear magnetorelaxometry,” *Physics in Medicine & Biology*, vol. 68, p. 155002, 2023. doi: 10.1088/1361-6560/ace304.

 DOI 10.15203/99106-194-6-11

# Synthetic Biosignal Generation for Wearable Device Testing

M. Ferencz, R. Drath, B. Dömer

***Abstract*—Sleep-disordered breathing, including obstructive and central sleep apnea, carries significant health risks. Validating and benchmarking wearable devices for its detection typically relies on complex and resource-intensive clinical studies. To reduce dependence on such studies, simulation and bench-test systems may be used, but they require realistic, yet parametrizable physiological datasets. This paper proposes a concept for generating synthetic biosignals: electrocardiogram (ECG), photoplethysmogram (PPG), and respiratory waveforms together with inter-signal interactions such as respiratory-induced heart-rate variation and amplitude modulation of ECG and PPG signals. Two exemplary scenarios, obstructive and central apnea, demonstrate the framework's ability to reproduce clinically relevant patterns, including airflow reduction and blood oxygen desaturation. This software-based signal generator thus provides a configurable building block for simulation environments, enabling rapid and reproducible testing of wearable devices within controlled conditions and reducing the need for early-stage human trials.**

***Index Terms*—synthetic biosignals, electrocardiogram (ECG), photoplethysmography (PPG), sleep apnea simulation, signal generation, wearable device validation**

## I. Introduction

Sleep disorders such as obstructive sleep apnea (OSA) affect millions of individuals worldwide and are closely linked to cardiovascular complications. Diagnosis and monitoring typically rely on wearable medical devices that capture biosignals related to respiration, cardiac activity, and body movement. Traditionally, the validation of these devices requires clinical trials, which are time-consuming, costly, and subject to ethical constraints. Moreover, human-based studies typically fail to represent the full spectrum of real-world conditions, and the rapid pace of device innovation can render trial results outdated by the time they are published. Synthetic biosignal generation for selected scenarios, in combination with simulation hardware, offers a promising alternative for algorithmic validation and device testing without the mentioned issues.

Although several approaches exist for modeling biosignals, none offers a comprehensive multi-signal framework capable of generating all required waveforms including photoplethysmography (PPG), electrocardiography (ECG) and respiratory signals. Here, we mention the comprehensive model of PNEUMA [1], which simulates physiological processes but does not generate PPG or ECG waveforms. The Pulse Physiology Engine (PPE) [2] focuses on producing PPG, ECG and respiratory signals with broadly universal morphology, but it does not capture the diversity of biosignals and their physiological interactions. Other approaches offer more direct control over specific waveforms, such as the one by McSharry et al. [3] for ECG generation, and by Martín-Martínez et al. [4] for PPG generation. However, these methods are designed for single-signal synthesis and therefore fail to represent the coupled interplay of physiological processes. An overview of existing approaches is provided in Table I.

TABLE I: Summary of existing approaches for generating biosignals

| Framework | PPG | ECG | Respiration | Interactions |
|---|---|---|---|---|
| PNEUMA [1] | ✗ | ✗ | ✓ | ✓ |
| PPE [2] | ✓ | ✓ | ✓ | ✗ |
| McSharry et al. [3] | ✗ | ✓ | ✗ | ✗ |
| Martín-Martínez et al. [4] | ✓ | ✗ | ✗ | ✗ |
| Proposed approach | ✓ | ✓ | ✓ | ✓ |

To address these limitations, the objective of this work is to develop a signal-generation framework that enables the definition, parameterization, and reproducible generation of synthetic physiological signals for testing and validation. The framework is designed to generate PPG, ECG and respiratory waveforms in a physiologically consistent manner. It should further integrate models of inter-signal dependencies, thereby enabling realistic simulations of sleep-disordered breathing scenarios, such as obstructive and central sleep apnea.

## II. Methods

### A. Signal Generation Framework

The proposed concept for signal-generation is implemented as a software framework using Python (3.12) and the NumPy library, with plans to release it as an open-source code. It contains the following main modules:

*1) ECG:* Heartbeat activity is modeled as spike trains based on user-defined heart rate and heart rate variability. RR intervals follow a Gaussian distribution constrained by physiological limits. ECG morphology is synthesized by convolving the spike train with Gaussian prototypes representing the P, Q, R, S, and T waves, similar to the approach presented in [3].

*2) PPG:* Pulse oximetry relies on PPG signals generated by the absorption characteristics of tissue at two wavelengths, red (R) and infrared (IR). The resulting transmitted or reflected light is sensed by a photodiode (PD), which converts it into an electrical current. The generation of PPG involves the following two aspects: a) Amplitude: AC and DC components of red and infrared PPGs are calculated based on any four of the following parameters, defined by the user: $AC_{Red}$, $DC_{Red}$, $AC_{IR}$, $DC_{IR}$, Perfusion Index ($PI_{IR}$) and blood oxygen saturation ($SpO_2$). b) Waveform: Each PPG pulse wave is generated using first derivative fiducial points that were derived from PPG recordings of real patients. Imposing constraints that

M. Ferencz is with Pforzheim University (75175 Pforzheim, Tiefenbronner Str. 65, Germany) and Löwenstein Medical Technology GmbH & Co. KG (76135 Karlsruhe, Südendstr. 42, Germany), correspondence e-mail: mate.ferencz@hs-pforzheim.de.

R. Drath and B. Dömer are with Pforzheim University.

enforce unit amplitude of the pulse wave and zero first derivatives at the beginning and end of each pulse ensures normalization and enables smooth concatenation of successive pulses without discontinuities. The pulse wave is subsequently scaled using the AC amplitude and combined with the DC component, producing the final signal.

*3) Respiratory Signals:* Respiratory mechanics are modeled using the single-compartment lung model [5] with resistance (R) and compliance (C) parameters, coupled with the effort model of Fresnel et al. [6], parametrized by ventilatory rate ($f_v$) and occlusion pressure ($P_{0.1}$). Volume and flow signals are computed using the analytical solution of the differential equation governing lung and effort dynamics.

### B. Signal Interactions

To model the physiological interaction between the generated signals, three phenomena were chosen:

*1) Respiratory Sinus Arrhythmia (RSA):* Heart rate is known to increase during inspiration and decrease during expiration. This effect is modeled by applying a heart-rate bias directly proportional to the lung volume.

*2) Amplitude Modulation:* A lung volume proportional variation is applied to the ECG baseline, the R-wave amplitude, PPG perfusion index, and PPG DC amplitude while strictly preserving $SpO_2$ levels.

*3) Pulse Transit Time (PTT):* The time delay between heartbeats extracted from the ECG and PPG signal, known as PTT, is of clinical importance. Higher blood pressure corresponds to a shorter, whereas lower blood pressure results in a longer PTT. In our model, this parameter can be explicitly set using the fiducial points determining the delay between the ECG R-wave and the pulse wave upstroke.

### C. Exemplary scenarios

Two scenarios were configured to illustrate the framework's capabilities: obstructive sleep apnea (OSA) and central sleep apnea (CSA). Model parameters were set empirically to generate waveforms similar to polysomnography recordings.

*1) OSA:* In the OSA scenario, an increased airway resistance reduces airflow, while respiratory drive is maintained. $SpO_2$ declines until an arousal restores airway patency, followed by rapid oxygen saturation recovery.

*2) CSA:* In the CSA scenario, respiratory drive diminishes while airway resistance remains constant. This results in a gradual decrease and increase of $SpO_2$. In contrast to OSA, respiratory drive and oxygen saturation recovers slower.

## III. Results

Fig. 1 and 2 show the synthetically generated signals for the OSA and CSA scenario. In both cases, we observe an increase in $AC_{\mathrm{Red}}$, corresponding to a desaturation during apneas.

In OSA, respiratory flow decreases despite maintained effort, whereas in CSA the reduction in flow arises from diminished effort. Differences are also visible in resaturation dynamics. Furthermore, the OSA-related PPG signal shows greater respiratory induced amplitude modulation and a pulse waveform with a distinct dicrotic notch, characteristic of the younger OSA patient group.

## IV. Discussion

The proposed framework successfully generates synthetic biosignals replicating key patterns of sleep-disordered breathing. By modeling ECG, PPG, and respiratory signals in an integrated manner, the system captures inter-signal dependencies, which are essential for realistic simulation.

We note that simplifications were applied in this study. Respiratory influence on ECG and PPG signals was modeled using a simplistic relationship, and blood oxygen desaturation dynamics were predefined rather than derived from a blood gas model incorporating airflow.

Potential improvements include a more realistic model of respiration-related variations in ECG and PPG signals, as well as implementing a physiologically based model in which oxygen desaturation is connected to respiratory flow and blood-gas exchange. Coupling this with arousal mechanisms that trigger changes in heart rate and breathing would further enhance the realism of the generated signals.

In addition, modeling movement and the resulting artifacts in ECG and PPG signals would be a valuable extension. To accurately parameterize and validate these models, sufficiently large datasets containing real patient data would be required.

These enhancements would improve the realism of synthetic datasets, enabling more robust validation of medical devices without the constraints of clinical trials.

## V. Conclusion

This work presents a concept and a framework for generating synthetic physiological signals, including ECG, PPG, and respiratory waveforms. By simulating obstructive and central sleep apnea scenarios, it demonstrates its potential for preclinical device testing. The ability to model signal interactions and adjust physiological parameters makes this approach highly adaptable for a wide range of research and development applications in sleep medicine and wearable technology. Future enhancements will focus on incorporating more physiologically accurate models of oxygen desaturation, arousal mechanisms, and signal artifacts to further enhance realism and clinical relevance.

## Acknowledgements

This work was funded by the BMFTR and the federal state of Baden-Württemberg within the program "Forschung an HAW" under grant number 13FH532KX2.

## References

[1] H. Fan and M.C.K. Khoo. Pneuma - a comprehensive cardiorespiratory model. pages 1533–1534 vol.2, 2002.

[2] A. Bray, J. B. Webb, A. Enquobahrie, J. Vicory, J. Heneghan, R. Hubal, S. TerMaath, P. Asare, and R. B. Clipp. Pulse physiology engine: an open-source software platform for computational modeling of human medical simulation. *SN Comprehensive Clinical Medicine*, 1(5):362–377, 2019.

[3] P.E. McSharry, G.D. Clifford, L. Tarassenko, and L.A. Smith. A dynamical model for generating synthetic electrocardiogram signals. *IEEE Transactions on Biomedical Engineering*, 50(3):289–294, 2003.

[4] D. Martin-Martinez, P. Casaseca-de-la Higuera, M. Martin-Fernandez, and C. Alberola-Lopez. Stochastic modeling of the ppg signal: A synthesis-by-analysis approach with applications. *IEEE Transactions on Biomedical Engineering*, 60(9):2432–2441, 2013.

[5] J. H. T. Bates. Lung mechanics: An inverse modeling approach. 2009.

[6] E. Fresnel, J. Muir, and C. Letellier. Realistic human muscle pressure for driving a mechanical lung. *EPJ Nonlinear Biomedical Physics*, 2(1), 2014.

 DOI 10.15203/99106-194-6-11

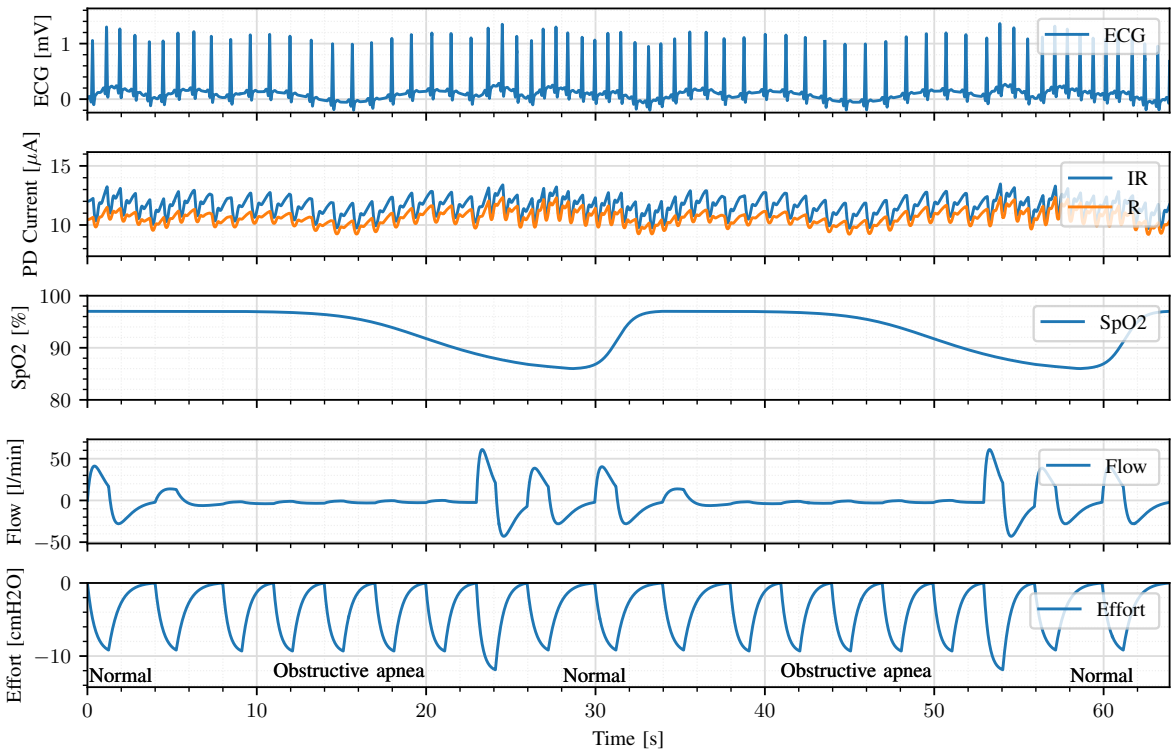


Fig. 1: Synthetically generated signals illustrating the OSA scenario, including ECG, PPG, $SpO_2$, respiratory flow and effort.

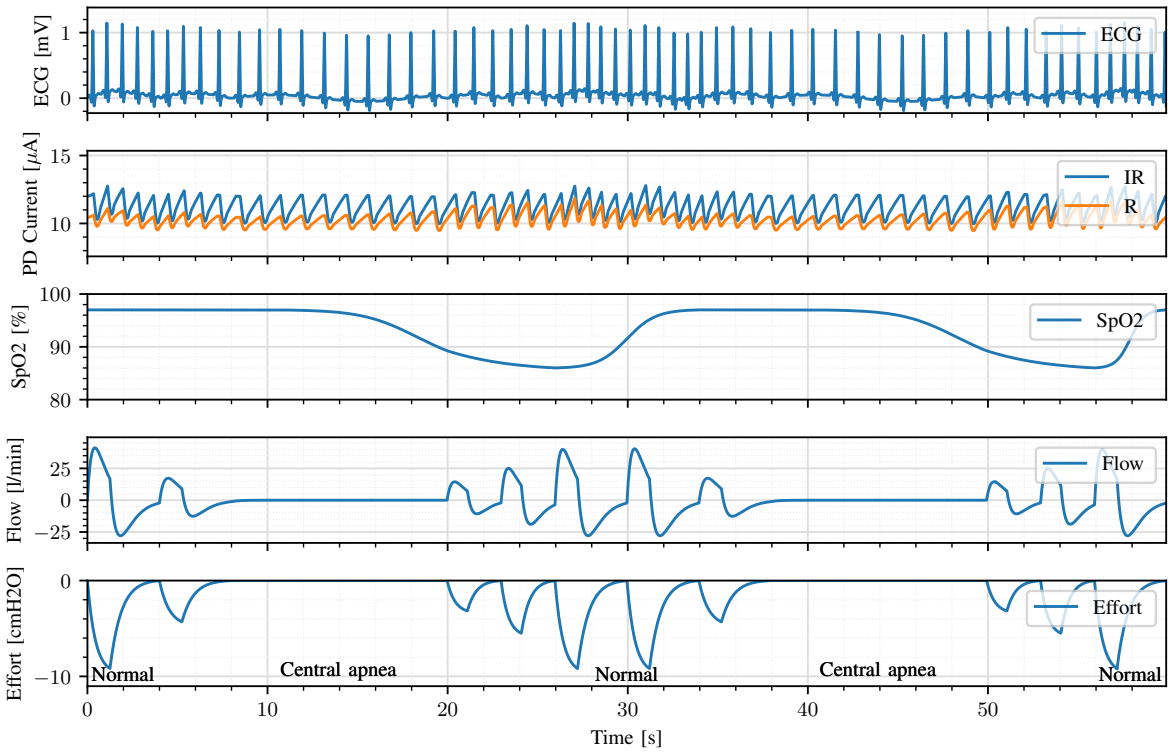


Fig. 2: Synthetically generated signals illustrating the CSA, including ECG, PPG, $SpO_2$, respiratory flow and effort.

 DOI 10.15203/99106-194-6-12

# EEG Signatures of Skin-to-Skin Touch in Romantic Couples

C. Schranz, A. Schirmer

***Abstract*—Much research attests to the pleasantness of gentle touch and implicates C-tactile (CT) afferents, a special class of mechanoreceptors, in this effect. However, we currently lack insights from skin-to-skin contact in natural social interactions. We addressed this gap in an EEG/ERP study examining touch in romantic couples. We found that early somatosensory signaling was enhanced for CT-absent relative to CT-present touch. By contrast, a later component thought to index CT processes was greater for CT-present than for CT absent touch. These effects compared for touch giving and receiving and dissociated from touch pleasantness indicating that the latter arises from multiple processes only some of which are CT-relevant.**

***Index Terms*—social touch, EEG, couples, pleasantness**

## I. Introduction

It has long been known that physical contact with another individual can improve affect and wellbeing. Moreover, research across the past two decades has revealed aspects of the mechanistic pathway underpinning these effects, including a special somatosensory receptor referred to as C-tactile (CT) afferent situated in hairy skin [1], [2]. The properties of this receptor suggest a particular relevance for gentle skin-to-skin contact. Unlike other mechanoreceptors, CTs are preferentially activated by gentle, dynamic contact with skin temperature [3]. CT input to the brain reaches the insula as well as socio-affective circuits [4]. It has been linked to neural correlates such as an attenuation in theta activity at multiple scalp sites [6] and somatosensory event-related potentials (ERPs) [7]. Moreover, CT activity predicts tactile pleasure and appears to be stress-buffering [8]. CT touch has also been shown to be pain-dampening since it attenuates both subjective pain ratings as well as laser evoked potentials (N1 and N2-P2 potentials) [9].

Currently, our understanding of how CT input is represented in the brain is still incomplete. Past neuroimaging studies have relied on artificial touch stimuli delivered by a robot or human holding a brush instead of using natural skin-to-skin touch. This is because measuring time-locked brain responses to skin-to-skin touch is technically challenging both in terms of obtaining precise touch onset information and controlling the physical properties of touch. Here, we tackled these challenges wishing to compare ERPs elicited to CT relevant and CT irrelevant skin-to-skin touch.

## II. Methods

Towards this end, we developed an EEG paradigm for romantic couples (data collection ongoing) in which partners are seated diagonally across from each other and asked to place their forearm on a table positioned between them. Curtains prevent the two partners from seeing each other and a monitor in front of each of them provides instructions on a trial-by-trial basis. At the beginning of each trial, participants are informed whether they should gently stroke or be gently stroked by their partner and whether they should make physical contact using hairy, CT-innervated skin or glabrous skin without CTs. A galvanometer developed in our lab with electrodes attached to each partner records touch onset and sends corresponding temporal markers to the EEG. After two seconds of touching, participants are prompted to stop touching and to rate touch pleasantness.

The pre-registered sample size for this study is 35 couples. Here, we present a preliminary report of the first 19 couples that visited our lab. Their data was recorded at 500 Hz with a 64 channel ANT EEGo system. CPz was used as the online reference. Electrode impedance was below 20 kΩ. EEG data was filtered using a 0.5-30 Hz bandpass filter and epoched from -1s before touch initiation to 1s after. Artifacts were removed using visual inspection and independent component analysis. A current source densitiy transformation was applied for data rereferencing. The data was baseline corrected from -200 to 0 ms relative to touch onsets. Our ERP analysis has been focused on the somatosensory potentials evoked over somatosensory cortex contralateral to the stimulation arm at channels CP1 and CP3.

## III. Results

Both touch giving and touch receiving were perceived with similar pleasantness however stroking with glabrous skin was perceived as more pleasant than stroking with hairy skin for those providing the stroking. Those receiving the stroking preferred being touched on hairy skin (Fig. 1.).

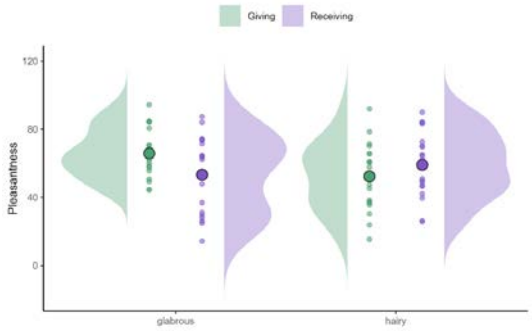


Fig. 1: Pleasantness by skin types and roles with means highlighted

In the ERP, an early negativity peaking 70 ms following touch onset was larger when touch was perceived with glabrous than with hairy skin, in line with the fact that this skin type is more densely innervated by mechanoreceptors in general and creates more salient cortical representations. Interestingly, however,

C. Schranz and A. Schirmer are with the Department of General Psychology II, University of Innsbruck, Austria, e-mail: christian.schranz@uibk.ac.at.

the sN400, a later negativity peaking between 300 and 400 ms, was larger for CT-innervated hairy than for CT-free glabrous skin irrespective of whether participants were giving or receiving touch (Fig. 2.).

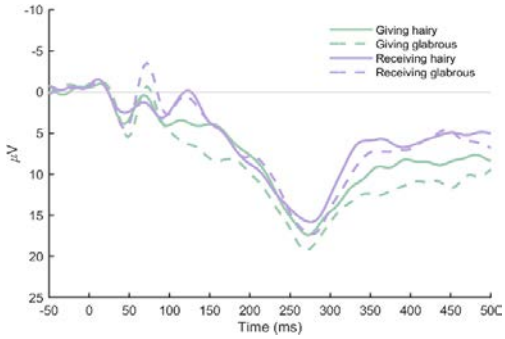


Fig. 2.: ROI ERPs by roles and skin type

This agrees with previous evidence of larger sN400 amplitudes to brushstrokes at CT optimal when compared with suboptimal stroking and a positive relationship between this negativity and the pleasantness of touch [7], [10].

When providing the stroking, somatosensory potentials were generally suppressed, as would be expected from corollary discharge. Interestingly, however, somatosensory suppression increased with time, especially for glabrous when compared with hairy skin, thus amplifying the differences between CT relevant and irrelevant touch observed for the sN400 during touch receiving.

## IV. Conclusions

In sum, we present the first somatosensory potentials elicited by skin-to-skin touch. Our results show that such potentials can be reliably measured, identifying components similar to those evoked by artificial touch stimuli. Moreover, they reveal original differences in tactile processing as a function of CT innervation starting 70 ms following touch onset. Most notably, the sN400, a late negative component, is larger for CT relevant touch and its CT response less attenuated by active touching underscoring the particular relevance CTs have in the context of human skin-to-skin touch. Notably, ERP results diverge from the pleasantness ratings indicating that these ratings result from multiple somatosensory processes but likely also from the contribution of proprioceptive and motor processes associated with the different touch positions. For example, the finger position and motion for glabrous stroking may have been more comfortable than for hairy stroking and this, rather than the somatosensory stimulation, may have dominated how a touch felt.

Together, the present study provides original evidence on the relevance of CT processes in natural skin-to-skin touch. Its technology development and results are relevant for the design of future touch research and touch-based interventions in naturalistic social settings.

## References

[1] H. Olausson *et al.*, ‘Unmyelinated tactile afferents signal touch and project to insular cortex’, *Nat. Neurosci.*, vol. 5, no. 9, pp. 900–904, Sep. 2002.

[2] Å. B. Vallbo, H. Olausson, and J. Wessberg, ‘Unmyelinated Afferents Constitute a Second System Coding Tactile Stimuli of the Human Hairy Skin’, *J. Neurophysiol.*, vol. 81, no. 6, pp. 2753–2763, Jan. 1999.

[3] R. Ackerley, H. Backlund Wasling, J. Liljencrantz, H. Olausson, R. D. Johnson, and J. Wessberg, ‘Human C-tactile afferents are tuned to the temperature of a skin-stroking caress’, *J. Neurosci.*, vol. 34, no. 8, pp. 2879–2883, Feb. 2014.

[4] A. Schirmer, F. McGlone, ‘A touching Sight: EEG/ERP correlates for the vicarious processing of affectionate touch’, *Cortex*. Feb. 111:1-15, 2019 .

[5] I. Morrison, L. S. Löken, J. Minde, J. Wessberg, I. Perini, I. Nennesmo, H. Olausson, ‚Reduced C-afferent fibre density affects perceived pleasantness and empathy for touch’ *Brain : a journal of neurology*,*134*(Pt 4), 1116–1126, 2011.

[6] M. von Mohr, M.J. Crowley, J. Walthall, L.C. Mayes, K.A. Pelphrey, H. J. V. Rutherford, ‘ EEG captures affective touch: CT-optimal touch and neural oscillations.’ *Cogn Affect Behav Neurosci* **18**, 155–166, 2018.

[7] A. Schirmer, O. Lai, F. McGlone, C. Cham, and D. Lau, ‘Gentle stroking elicits somatosensory ERP that differentiates between hairy and glabrous skin’, *Soc. Cogn. Affect. Neurosci.*, vol. 17, no. 9, pp. 864–875, Sep. 2022.

[8] M. von Mohr, C. Krahé, B. Beck, A. Fotopoulou, ‘The social buffering of pain by affective touch: a laser-evoked potential study in romantic couples’ *Soc Cogn Affect Neurosci*. Nov 8;13(11):1121-1130, 2018.

[9] D. Gursul *et al.*, ‘Stroking modulates noxious-evoked brain activity in human infants’, *Curr. Biol. CB*, vol. 28, no. 24, pp. R1380–R1381, Dec. 2018.

[10] A. Schirmer, O. Lai, C. Cham, and C. Lo, ‘Velocity-tuning of somatosensory EEG predicts the pleasantness of gentle caress’, *NeuroImage*, vol. 265, p. 119811, Jan. 2023.

.

 DOI 10.15203/99106-194-6-13

# Rapid Degaussing for Highly Sensitive Magnetometers: A Case Study of the Kiel Magnetically Shielded Room (KMSR)

J. Schütt*, E. Elzenheimer*, A. Schnabel *, J. Arbustini, H. Matz, J. Voigt, M. Höft

***Abstract*—With the commercialization of Spin-Exchange-Relaxation-Free Optically Pumped Magnetometers (SERF-OPMs), OPMs have emerged as a non-cryogenic alternative to Superconducting Quantum Interference Devices (SQUIDs) owing to their compactness and flexibility. However, maximizing SERF-OPM potential requires low residual fields and minimal gradients, necessitating a high-performance Magnetically Shielded Room (MSR). Since the shielding material (Mu-metal) of an MSR exhibits hysteretic behavior, a well-controlled degaussing procedure is required to return the material to a low-remanence state. This work presents the initial quantitative assessment of the degaussing performance of the Kiel MSR (KMSR) using two strategies for driving the built-in degaussing coils. In the first, degaussing is performed using a manually controlled $50\,\mathrm{Hz}$ rotary transformer supplied by the MSR manufacturer. In the second, a semi-automated approach is employed, leveraging additional hardware. The manual approach yields residual fields of $11\,\mathrm{nT}$ within the central $1\,\mathrm{m}^3$ volume, with pronounced spatial gradients and poor run-to-run reproducibility. The other approach reduces the residual magnitude to $4.7\,\mathrm{nT}$ and achieves a minimal field at the MSR center below $0.4\,\mathrm{nT}$. This approach, with a degaussing time of under 4 minutes, provides a low, stable environment for SERF-OPMs and offers a practical route to low-cost, fast, and user-friendly upgrades for existing MSRs with integrated coils.**

***Index Terms*—Magnetically Shielded Room, Degaussing, Mu-Metall, Magnetometers, Biomagnetism, Magnetic Fields, Hysteresis, Fluxgate, Optically Pumped Magnetometers**

## I. Introduction

The detection of biomagnetic signals has transitioned from the use of large, cryogenic sensor systems to compact and flexible magnetometers. Among these, Spin-Exchange Relaxation-Free Optically Pumped Magnetometers (SERF-OPMs) represent an advancement in the field. However, SERF-OPMs must operate under near-zero-field conditions to maintain reliable performance [1], owing to their underlying operational principle. The internal compensation coils incorporated into the OPM sensor head can counteract only a limited range of static residual fields [2]. As an OPM sensor is zeroed and initialized at a specific residual static field strength by these coils, temporal variations in the residual field can shift the operating point, thus significantly affecting the measured signal due to the intrinsic sensor system behavior. Some manufacturers, such as QuSpin, specify a required ambient field well below $50\,\mathrm{nT}$ compensation range or lower for their QZFM Gen-3 sensor, while its closed-loop operating range of only $\pm 5\,\mathrm{nT}$ indicates a much lower tolerance to field drift and gradients [3]. Similar constraints apply to multichannel SERF-OPM systems, such as those used in Magnetoencephalography (MEG), including systems like the FieldLine HEDscan or QuSpin Neuro-1 [4]. Furthermore, SERF-OPMs are susceptible to Cross-Axis Projection Error (CAPE), whereby the measured signal amplitude deviates from the actual field value due to an orthogonal quasi-static magnetic field component relative to the sensitive axis. Both field fluctuations and CAPE degrade measurement reproducibility, underscoring the need for a Magnetically Shielded Room (MSR) with sufficiently low remaining magnetic field and field gradients, which are achievable by a reliable degaussing procedure to enable OPM-based biomagnetic measurements. The degaussing process (more precisely: equilibration or idealization) is typically accomplished by applying a sinusoidal signal to the MSR built-in degaussing coils at a constant, very low frequency and varying amplitude. Thus, a coil current with gradually increasing amplitude drives the shielding material into saturation by generating a strong magnetic field. Subsequently, the coil current is gradually reduced so that the shielding material is idealized to the outer environmental field. At the Kiel MSR (KMSR), two different degaussing setups were investigated: a manually operated $50\,\mathrm{Hz}$ rotary transformer (classical $50\,\mathrm{Hz}$ VARIAC) and a semi-automated system with a computer-generated waveform driving a four-quadrant amplifier and isolation transformer. Both degaussing approaches are evaluated identically to assess the resulting residual field within the KMSR.

## II. Methods

The two degaussing approaches are evaluated by comparing the residual static (DC) field vectors, obtained from a 3D mapping in the central $1\,\mathrm{m}^3$ of the KMSR. This volume is sampled manually on a $5 \times 5 \times 5$ grid with $25\,\mathrm{cm}$ spacing between the measurement points. For this purpose, a three-axis fluxgate magnetometer (MAG03 MSL 70, Bartington, Witney, GB) was mounted on an aluminum positioning framework and manually moved to each location within the central $1\,\mathrm{m}^3$ volume by a person wearing non-magnetic clothing inside the KMSR. A full scan of the 125 points required approximately

*Authors contributed equally to this work. Corresponding author: E. Elzenheimer (e-mail: ee@tf.uni-kiel.de).

J. Schütt, E. Elzenheimer, J. Arbustini and M. Höft are with the Kiel University, Germany.

H. Matz, E. Elzenheimer (2nd-Afflination) and J. Voigt are with the Physikalisch-Technische Bundesanstalt (PTB), Berlin, Germany.

A. Schnabel is a former Senior Scientist from the PTB Berlin, Germany.

 DOI 10.15203/99106-194-6-13

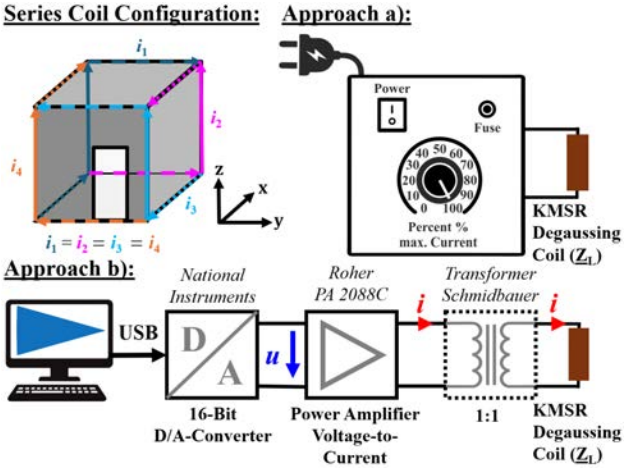


Fig. 1: Coil configuration and both available approaches.

25-30 minutes. The KMSR is a chamber built by Vacuumschmelze (Modified Ak3B, 3 layer + 1 Al, Vacuumschmelze, Hanau, DE) with a usable volume of $3\,\mathrm{m} \times 2.4\,\mathrm{m} \times 2.4\,\mathrm{m}$. The built-in degaussing coils, wound with five turns of $2.5\,\mathrm{mm}^2$ copper wire, encircle only the two inner shielding (Mu-metal) layers and are wired in a fixed configuration, as shown in Fig. 1. For fluxgate characterization regarding directivity, see [1]. **Approach a)** uses a manually operated transformer supplied by the MSR manufacturer (Vacuumschmelze, Hanau, DE), whereas **Approach b)** employs a semi-automated setup based on a DAQ system with D/A conversion card, a dedicated voltage–current power amplifier, and a low-frequency transformer with a transformation ratio of 1:1. Fig. 1 provides a schematic overview of both available degaussing approaches.

a) **Manually Operated Transformer:** A $50\,\mathrm{Hz}$ manually operated rotary transformer (VARIAC) is used. The VARIAC is connected to the main supply ($230\,\mathrm{V}/50\,\mathrm{Hz}$) and galvanically isolated to the built-in degaussing coils, through which it drives the required degaussing current. First, the coil current is increased using the transformer potentiometer to drive the Mu-metal into saturation. Subsequently, the current is gradually reduced to zero. The manual approach relies heavily on the operator and potentiometer settings, which are neither recoverable nor repeatable.

b) **Semi-Automated Procedure:** It uses equipment from the Physikalisch-Technische Bundesanstalt (PTB) Berlin, which was already applied successfully to the MSR at the Fraunhofer IPM in Freiburg [5] and employs a Rohrer four-quadrant voltage-current power amplifier (Rohrer PA2088C, Munich, DE) in combination with a $7\text{-}10\,\mathrm{Hz}$-isolation transformer built by Schmidbauer Transformatoren- und Gerätebau GmbH (Hebertsfelden, DE). The used degaussing waveform consists of a $7\,\mathrm{Hz}$ sinusoidal signal that is linearly ramped up over 10 periods, held at its maximum amplitude at $27.9\,\mathrm{A}$ for another 10 periods, and subsequently linearly ramped down over 1500 periods. The less than 4-minute-long sequence is generated by a LabVIEW control and converted into an analog voltage using a 16-bit analog output module by National Instruments (NI-DAQCard-6036, Austin, US). The amplifier converts the generated voltage signal into a current and feeds it through the isolated transformer, which acts as a high-pass filter to suppress any DC offset. The offset-free sinusoidal current waveform is then applied to the degaussing coils.

## III. Results

The KMSR, before any degaussing approach was performed, exhibited a maximum residual field magnitude of nearly $15\,\mathrm{nT}$ near the upper door area (Fig. 2, left). Degaussing with the manufacturer-supplied manual transformer yielded comparatively high residual fields of up to $11\,\mathrm{nT}$ (Fig. 2, middle), and further investigations revealed markedly poorer run-to-run reproducibility (not shown) than with the semi-automated approach. Applying the $7\,\mathrm{Hz}$, 10–10–1500 periods generated sine waveform with varying amplitude yielded reproducible low residual fields in the central $1\,\mathrm{m}^3$ volume (Fig. 2, right) and individual runs exhibited maxima in the range of $4.4\text{-}5.4\,\mathrm{nT}$. The average of five independent degaussing runs identified a maximum of $4.7\,\mathrm{nT}$, while the mean field magnitude at the room center remained below $0.4\,\mathrm{nT}$. All maps presented in Fig. 2 consistently highlighted the door region as the expected dominant source of non-uniformity (except for the right figure, where a structure in the top is dominant, which is explained in Sec. IV). Opening and closing the door altered the field distribution in a reproducible manner, independent of coil polarity, and locally reduced or even inverted the field near the door without re-degaussing. Extending the degaussing periods did not produce any significant change in the residual field. Since the repeatability of OPM measurements depends not only on a low static residual field but also on its temporal stability, a repeated measurement of the magnetic field at the center of the $1\,\mathrm{m}^3$ cube was performed. This measurement showed that the residual field varies by up to $\pm 1.5\,\mathrm{nT}$ over a period of $1.5\,\mathrm{h}$.

## IV. Discussion

The results from Sec. III demonstrate that the semi-automated degaussing approach, when applied to the KMSR, produces a low static residual magnetic field with minimal field gradients, suitable for SERF-OPM operation. With maximal residual field magnitudes in the range of $4.4\text{–}5.4\,\mathrm{nT}$ and MSR center-field values below $0.4\,\mathrm{nT}$, the residual field remains well below the $50\,\mathrm{nT}$ ambient-field requirement exemplary specified, for example, by QuSpin [3]. A vortex-like feature (see Fig. 2, right) near the ceiling is observed. Follow-up investigations with a handheld fluxgate (Fluxmaster, Stefan Mayer Instruments, Dinslaken, DE) indicate that the layout of the current-carrying lighting cables in the MSR also affects the residual field, locally magnetizing the inner Mu-metal layer. The residual field in the central $1\,\mathrm{m}^3$ volume could likely be further reduced if the current lighting in the MSR were replaced by a fiber-optic lighting system. The temporal stability of the degaussed state is sufficient for OPM-MEG

 DOI 10.15203/99106-194-6-13

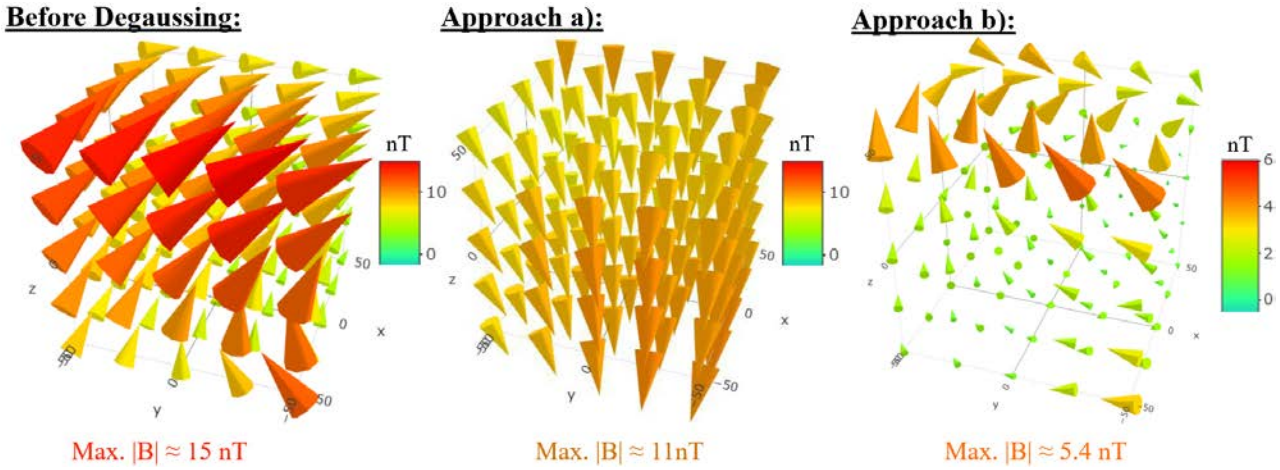


Fig. 2: Results of the 3D residual static (DC) field scan before degaussing and after applying approaches a) and b). Note the different color scaling for approach b).

applications. The measured central field varies only up to $\pm 1.5\,\mathrm{nT}$ over $1.5\,\mathrm{h}$. This variation is primarily attributed to fluxgate DC-offset drifts. The DC-offsets were determined before and after recording a field map (mean-value of the 180° rotation). The drift was assumed to be linear and subtracted before data analysis. The remaining predominant variations are attributable to external factors, such as opening and closing the building's front door or the movement of large metallic objects near the building (e.g., a car), rather than to inherent limitations of the degaussing process itself. These variations remain within the $\pm 5\,\mathrm{nT}$ operating range of QuSpin sensors, indicating that the KMSR provides a suitable environment for SERF-OPM operation.

## V. Conclusion

This case study quantitatively demonstrates that replacing the manually operated 50 Hz-transformer approach with a semi-automated degaussing procedure based on a four-quadrant amplifier and a 7-10 Hz-isolation transformer can substantially improve the performance of the KMSR. By suppressing DC offsets and common 50 Hz harmonics, this operator-independent setup achieves a level of reproducibility that the manual approach cannot provide. Although only the two inner Mu-metal shells of the KMSR are equipped with degaussing coils, the resulting static residual field distribution after a degaussing time period of less than 4 minutes meets the requirements of commonly used SERF-OPMs. The averaged distribution is characterized by maximum values below $5\,\mathrm{nT}$ within a $1\,\mathrm{m}^3$ volume and less than $0.4\,\mathrm{nT}$ at the MSR center, with sufficient temporal stability to support multichannel OPM measurements. The demonstrated performance establishes a validated reference baseline for a planned rapid, low-cost, and user-friendly degaussing concept based on commercially available hardware and one-button operation.

## VI. Outlook

The next step is to implement a rapid degaussing setup suitable for daily operation, thereby ensuring proper conditions for multichannel OPM systems (e.g., MEG systems such as FieldLine HEDscan or QuSpin Neuro-1) and building on the results demonstrated here. This approach will use a hardware platform such as the STEMlab 125-14 (Red Pitaya, Solkan, SI) for waveform generation and D/A conversion, together with an AE Techron amplifier (AE Techron 7224, Elkhart, US) and a 7-Hz-1kVA-isolation-transformer built by Schmidbauer (13336A, Hebertsfelden, DE). The aim is to use a standalone, low-cost SDR platform for waveform generation, removing the need for a costly PC-based DAQ system and allowing one-button execution of the full degaussing sequence within minutes. Such an approach is expected to enable frequent and reproducible degaussing routines for biomagnetic measurements in Kiel and to provide a practical upgrade path for existing MSRs with integrated coil systems.

## Acknowledgment

This work was funded by the DFG through project Z2 of CRC 1261 *"Magnetoelectric Sensors: From Composite Materials to Biomagnetic Diagnostics"* (Project ID 286471992).

## References

[1] E. Elzenheimer, S. Knappe-Grüneberg, J. Zerfowski, W. Evans, F. Grüneberg, M. Höft, S. R. Soekadar, S. Robinson, and J. Voigt, "Key Metrics and Experimental Test Bench for Assessing highly Sensitive Magnetometers in Research," *IEEE Sensors Journal*, vol. 25, no. 2, pp. 2432–2455, 2025.

[2] J. Osborne, J. Orton, O. Alem, and V. Shah, "Fully Integrated, Standalone Zero Field Optically Pumped Magnetometer for Biomagnetism," in *Steep Dispersion Engineering and Opto-Atomic Precision Metrology XI*, ser. Proc. SPIE, vol. 10548, 2018, pp. 89–95.

[3] QuSpin Inc., "QuSpin Zero-Field Magnetometer (QZFM)," https://quspin.com/products-qzfm/, accessed: Dec. 1, 2025.

[4] E. Elzenheimer, H. Matz, J. Zerfowski, P. Anders, M. Höft, R. Rieger, S. R. Soekadar, S. Robinson, and S. Knappe-Grüneberg, "Characterizing Timing Parameters in Commercial SERF-OPM Multichannel Systems for Biomagnetic Field Sensing," *arXiv preprint arXiv:2509.22198*, 2025.

[5] P. A. Koss, J. Voigt, R. Rasser, and A. Schnabel, "Fast Degaussing Procedure for a Magnetically Shielded Room," *Materials*, vol. 17, no. 23, p. 5877, 2024.

 DOI 10.15203/99106-194-6-14

# Testing a non-invasive ECG-based glucose classification for neonatal intensive care

C. Oprea, N. Hillbrand, L. Olivier, M. Schoberer, A. Stollenwerk

***Abstract*—Regulating glucose levels is crucial for neonates in intensive care units. Non-invasive continuous monitoring methods have been tested in adults, but not yet in neonates. This study explores an approach to glucose classification using electrocardiogram (ECG) signals, drawing on methodologies previously established for adult populations. The results demonstrate that change in glucose is easier to classify than the value itself and that classifiers trained on individual patients can achieve good accuracy and F1-scores. However, notable obstacles remain, including high inter-patient variability and the low occurrence of hypo- and hyperglycemic events, which hinder the development of a standardized algorithm. Future work should focus on addressing these challenges through personalized algorithms and data augmentation strategies to enhance the reliability of ECG-based continuous glucose monitoring in neonates.**

***Index Terms*—glucose monitoring, ECG, neonates**

## I. Introduction

Continuous monitoring of blood glucose is of high relevance for neonates in intensive care units. Hypoglycemic episodes are related to neurodevelopmental and cognitive impairment [1], while hyperglycemic episodes associate with neurological delay [2]. Furthermore, alternations between hypo- and hyperglycemia in very low birth weight infants are associated with increased mortality [3]. Blood gas analyses (BGA) can be used to obtain blood glucose values, but only offer spot measurements and cause patient stress and blood loss. Continuous glucose monitoring (CGM) is a technique enabling a higher measurement resolution, but is invasive and still in the testing phase for newborns [4]. The effect of different glucose levels on electrocardiogram (ECG) was evaluated in adults [5], paving the way for non-invasive hypo- and hyperglycemia detection approaches using ECG [6], [7] or ECG and accelerometer data [8]. From the ECG signal both heartbeat and heart variability features were extracted and used as input to machine learning algorithms, achieving good performance. However, the approach has not yet been applied to neonates. This paper presents the implemented pipeline for ECG delineation (the identification of fiducial points), feature extraction and glucose classification using BGA measurements as ground truth. The pipeline was used for two classification tasks, once to classify hypo-, normo- and hyperglycemia and once to classify the change in glucose as decreased, no change or increased. Each task was accomplished with three classifiers (logistic regression, random forest, neural network) and tested on multi- and single-patient datasets.

C. Oprea, N. Hillbrand, A Stollenwerk are with the Chair for Embedded Software (Informatik 11), RWTH Aachen University, Germany, e-mail: oprea@embedded.rwth-aachen.de

L. Olivier, M. Schoberer are with the Neonatology Section of the Department of Paediatric and Adolescent Medicine, RWTH Aachen University Hospital, Germany

## II. Methods

From an existing neonatal dataset, patients with BGA probes containing a blood glucose measurement were extracted together with the corresponding three-lead ECG recordings ($512\,\mathrm{Hz}$). Subsequently, 120 patients with a total of 2,341 BGA measurements were included in this study. From the three lead ECG measurements, only one lead is available as a signal. The specific lead that was available varied both within and between patients and there was no reliable annotation indicating the type of lead. Based on published CGM recordings in preterm neonates showing glucose changes of $\pm 1\,\mathrm{mg/dL}$ per minute, in extreme cases of $\pm 3\,\mathrm{mg/dL}$ per minute [9], [10], three minutes should suffice to capture a reliable glucose value. We added further 5 minutes to assure the ECG recordings are representative for the respective blood sample, due to known delays from the moment the sample was taken to the BGA's timestamp in the dataset. Thus, from the ECG data 8 minute segments were extracted before each annotated BGA timestamp. The data segments were then preprocessed to reduce noise and artifacts. First, absoute values above $3\,\mathrm{mV}$ were considered outliers and excluded. Next, to suppress baseline drift, a Butterworth bandpass filter of order 4 and a frequency range of $0.7\,\mathrm{Hz}$ to $70\,\mathrm{Hz}$ was applied. Lastly, a Savitzky-Golay smoothing filter of order 3 and window length $50\,\mathrm{ms}$ reduced high-frequency noise, while preserving the morphology of the ECG recording. For both filters the Python scipy signal library was used. The single preprocessing steps and their effect on the data are shown in Fig. 1.

The cleaned 8 minute long ECG signals were split into 80-beat windows to compute heartbeat, heart rate and variability features per window. The first step was to detect the single heartbeats and their fiducial points $P$, $Q$, $R$, $S$ and $T$ as well as the on- and offset points for the $QRS$ complexes and $P$- and $T$-waves. A commonly used delineation algorithm (NeuroKit2 toolbox [11]), did not perform well on our dataset, often missing the P-waves. Therefore, we developed our own algorithm. First, preliminary $R$-peaks and corresponding $QRS$ complexes were detected. Then the $R$-peaks were refined (due to different $QRS$ complexes such as $rR'$). Finally, $T$- and $P$-waves were detected as local maxima between $R$-peaks. Per heartbeat we extracted following features: $PR$, $QTc$, $RTc$ and $TpTe$, which have been shown to be impacted by hypo- and hyperglycemia [12]. These were averaged over the 80-beat windows to provide more robust measures than

 DOI 10.15203/99106-194-6-14

single beat computations. Further extracted features are: heart rate, standard deviation of $RR$ intervals, root mean square of successive $RR$ differences and percentage of successive $RR$ differences above $50\,\mathrm{ms}$. Using the intervals and heart variability features, additional extremity features were introduced to capture situations where ECG characteristics enter abnormal or physiologically stressed ranges. For each feature, a plausibility range was defined. Extremity was then calculated as the distance to the center of the range. This produced a measure indicating how far each feature was towards the extremes of its expected physiological range. Thus, the total feature set, computed over 80-beat windows, comprised sixteen features: four heartbeat-based features with corresponding four extremity features and four heart variability features with corresponding four extremity features. The number of features per 8 minute ECG recording varied, due to varying heart rates.

To learn the correspondence between ECG-based features and glucose, two classification tasks were defined:

**1. Glucose classification.** Classify ECG features into hypo-, normo- and hyperglycemia defined as: below $47\,\mathrm{md/dL}$, between $47\,\mathrm{md/dL}$ and $150\,\mathrm{md/dL}$ and above $150\,\mathrm{mg/dL}$, respectively. The input features are the 16 ECG features computed over 80-beat windows for each 8 minute ECG recording and the label is the blood glucose class, computed from the BGA glucose measurement.

**2. Glucose change classification.** Classify ECG features into three classes predicting the change in glucose between two consecutive BGAs. The classes are defined as $\Delta Glucose > -10\mathrm{mg/dL}$ (decrease), $|\Delta Glucose| < 10\mathrm{mg/dL}$ (no change) and $\Delta Glucose > 10\mathrm{mg/dL}$ (increase). The input features are the 16 ECG features computed over 80-beat windows for each 8 minute ECG recordings together with the last available glucose value. The label is the glucose change class, computed from two consecutive BGA glucose measurements.

The second task was introduced, to observe if glucose changes are easier to predict using information on the last available glucose measurement. The time elapsed between two BGAs is not constant but may vary depending on the patient and their condition (half an hour to multiple hours). A further reason for this task is based on research highlighting the risk of variations in blood glucose [3]. The classifiers were trained on the available multi-patient dataset but also on selected single-patient datasets. For each case the datasets were further split into train, validation and test with a 70-15-15 ratio. For the multi-patient dataset the test set contained only unseen patients and for the single-patient dataset we assured that the 80-beat windows from the same 8 minute ECG recording would be sorted into the same split, to avoid data leakage. Three models were implemented using the Python sklearn library: a logistic regression (LR) with maximal 1000 iterations, a random forest (RF) with 100 estimators and a neural network (NN): multi-layer perceptron with two hidden layers of 100 and 50 units respectively and maximal 500 iterations.

For the first task, due to severe class imbalance (hypoglycemia: 68, normoglycemia: 1836, hyperglycemia: 437), SMOTE [13] was employed to handle the imbalance. For logistic regression and neural networks additional class weights were set inversely proportional to class frequency.

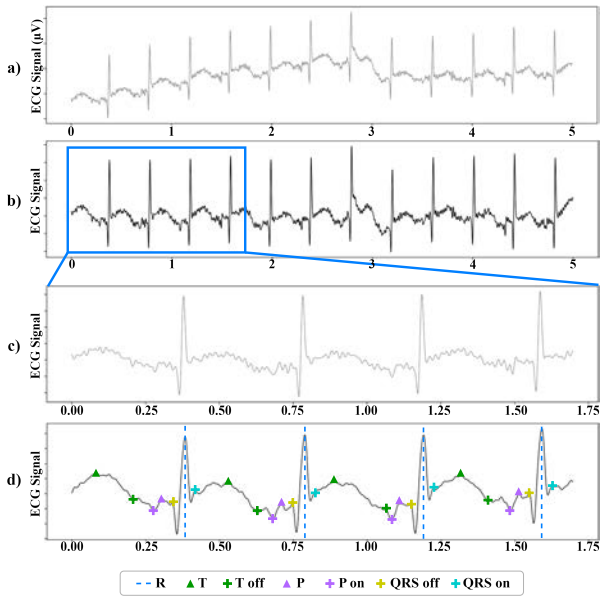


Fig. 1: Preprocessing of ECG signal: **a)** the original signal, **b)** signal filtered with Butterworth bandpass, **c)** extract of signal b), **d)** signal further filtered with Savitzky-Golay smoothing and delineated to identify fiducial points.

## III. Results

In Tab. I the results for the glucose classification task on the multi-patient dataset are presented. Overall the three algorithms achieved a low accuracy and F1 scores, with no clear superior across the three algorithms. The random forest performed best for the normoglycemic class, while the logistic regression and neural network achieved the best results for hypo- and hyperglycemia. As expected the F1 score was highest for the predominant normoglycemic class and lowest for the smallest class (hypoglycemia). The classification task on the single-patient datasets achieved for most patients better performance than on the multi-patient dataset. The patients with a more balanced fraction of hypo- and hyperglycemic samples, achieved better overall F1 scores. All three algorithms performed similarly good across patients. The results for the patient with the most glucose measurements (471) are shown in Tab. II. F1 score for hypoglycemia was not computed as no hypoglycemic measurements existed for this patient and thus no input was labeled as hypoglycemia.

The second task on glucose change classification achieved overall better results, with an increased accuracy but low F1 scores for the classes of an increased or decreased glucose measurement, again due to an imbalanced dataset. For this task, the random forest achieved the best results.

## IV. Discussion

The glucose classification task was setup as a three-class problem, to avoid learning a more concrete glucose value, for which the data would not have been sufficient and too sparsely spread. The results show, that even with the three class binning and commonly employed balancing techniques,

TABLE I: Results for task 1 on glucose classification for multi-patient dataset.

| Model | Accuracy | F1 (hypo) | F1 (normal) | F1 (hyper) |
|---|---|---|---|---|
| LR | 0.46 | **0.02** | 0.61 | **0.29** |
| RF | **0.57** | 0.00 | **0.72** | 0.06 |
| NN | 0.52 | 0.01 | 0.66 | 0.22 |

TABLE II: Results for task 1 on glucose classification for single-patient dataset, on patient with most BGA measurements (471).

| Model | Accuracy | F1 (hypo) | F1 (normal) | F1 (hyper) |
|---|---|---|---|---|
| LR | 0.75 | - | 0.79 | 0.71 |
| RF | 0.76 | - | 0.78 | 0.75 |
| NN | **0.78** | - | **0.79** | **0.76** |

TABLE III: Results for task 2 on glucose change classification for the multi-patient dataset. Increases and decreases by $10\,mg/dL$ were predicted.

| Model | Accuracy | F1 (dec.) | F1 (no change) | F1 (inc.) |
|---|---|---|---|---|
| LR | 0.84 | 0.27 | 0.93 | 0.22 |
| RF | **0.90** | **0.33** | **0.96** | **0.28** |
| NN | 0.89 | 0.22 | 0.95 | 0.21 |

the severe class imbalance was not overcome. Especially the hypoglycemic episodes were often misclassified. More promising results were obtained on the single-patient datasets, one being shown exemplarily. This result implies that a more balanced dataset could enable classifiers to learn features that distinguish the glucose classes. The second task, glucose change classification, achieved higher scores than the first task. The additional information on the last available glucose value and the task reformulation to classify the change could prove to be a viable path to extract information from ECG recordings on blood glucose. As blood gas analyses are a routine diagnostic tool on intensive care units, the available glucose measurements could be used as ground truth in a calibration step in future algorithms. Compared to the controlled setting of Tobore et al. [5], where the adults were sitting, our data indicates to be more difficult to process. The ECG recordings, even though filtered and smoothed, were upon investigation still very noisy (e.g. due to movement artifacts), which heavily influenced the delineation quality. Furthermore, the morphology of the recordings and the heart rate variability can be impacted through various mechanisms. To overcome this variability, further analysis and grouping of the ECG signal could be employed, to correct for selected mechanisms and varying ECG leads. Another strategy would be to create personalized classifiers based on groups of patients with similar ECG morphologies.

A limitation of our work is the delineation algorithm, which, despite being adapted to our data, still has problems with correctly identifying the fiducial points. A further drawback was the availability of only spot measurements of glucose to use as labels, which restricted the data set and did not reflect the continuous change of glucose.

## V. Conclusion

A pipeline was developed to preprocess neonatal ECG recordings, extract features and train and test glucose classification through two different task setups and evaluation datasets. The pipeline was developed leaning on existing work for adults, to test its transferability to neonatal intensive care data. Classifying ECG to predict change in glucose proved to be more efficient than directly classifying into hypo-, normo- and hyperglycemia. The training on single-patient datasets with more balanced classes improved the quality of the classifiers performance, proving the feasibility of using ECG to estimate glucose. Through future work, the preprocessing and classification could be further analyzed and improved.

## Acknowledgment

Research funding: Federal Ministry of Research, Technology and Space (031L0303A). Computations were performed with computing resources granted by RWTH Aachen University under project thes2043.

## References

[1] S. Diggikar et al., “Neonatal Hypoglycemia and Neurodevelopmental Outcomes—An Updated Systematic Review and Meta-Analysis,” Life, vol. 14, no. 12, p. 1618, Dec. 2024, doi: 10.3390/life14121618.

[2] S. Guiducci, L. Meggiolaro, A. Righetto, M. Piccoli, E. Baraldi, and A. Galderisi, “Neonatal Hyperglycemia and Neurodevelopmental Outcomes in Preterm Infants: A Review,” Children, vol. 9, no. 10, p. 1541, Oct. 2022, doi: 10.3390/children9101541.

[3] D. Angelis, M. A. Jaleel, and L. P. Brion, “Hyperglycemia and prematurity: a narrative review,” Pediatr Res, vol. 94, no. 3, pp. 892–903, Sept. 2023, doi: 10.1038/s41390-023-02628-9.

[4] K. Beardsall et al., “Real-time continuous glucose monitoring in preterm infants (REACT): an international, open-label, randomised controlled trial,” Lancet Child Adolesc Health, vol. 5, no. 4, pp. 265–273, Apr. 2021, doi: 10.1016/S2352-4642(20)30367-9.

[5] I. Tobore, J. Li, A. Kandwal, L. Yuhang, Z. Nie, and L. Wang, “Statistical and spectral analysis of ECG signal towards achieving non-invasive blood glucose monitoring,” BMC Med Inform Decis Mak, vol. 19, no. 6, p. 266, Dec. 2019, doi: 10.1186/s12911-019-0959-9.

[6] L. L. Nguyen, S. Su, and H. T. Nguyen, “Neural network approach for non-invasive detection of hyperglycemia using electrocardiographic signals,” in 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Aug. 2014, pp. 4475–4478. doi: 10.1109/EMBC.2014.6944617.

[7] R. Cordeiro, N. Karimian, and Y. Park, “Hyperglycemia Identification Using ECG in Deep Learning Era,” Sensors, vol. 21, no. 18, p. 6263, Jan. 2021, doi: 10.3390/s21186263.

[8] D. Dave et al., “Detection of Hypoglycemia and Hyperglycemia Using Noninvasive Wearable Sensors: Electrocardiograms and Accelerometry,” J Diabetes Sci Technol, vol. 18, no. 2, pp. 351–362, Mar. 2024, doi: 10.1177/19322968221116393.

[9] A. Galderisi et al., “Continuous Glucose Monitoring in Very Preterm Infants: A Randomized Controlled Trial,” Pediatrics, vol. 140, no. 4, p. e20171162, Oct. 2017, doi: 10.1542/peds.2017-1162.

[10] A. Galderisi, M. Bruschettini, C. Russo, R. Hall, and D. Trevisanuto, “Continuous glucose monitoring for the prevention of morbidity and mortality in preterm infants,” Cochrane Database of Systematic Reviews, no. 12, 2020, doi: 10.1002/14651858.CD013309.pub2.

[11] D. Makowski et al., “NeuroKit2: A Python toolbox for neurophysiological signal processing,” Behav Res, vol. 53, no. 4, pp. 1689–1696, Aug. 2021, doi: 10.3758/s13428-020-01516-y.

[12] L. L. Nguyen, S. Su, and H. T. Nguyen, “Identification of Hypoglycemia and Hyperglycemia in Type 1 Diabetic patients using ECG parameters,” in 2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Aug. 2012, pp. 2716–2719. doi: 10.1109/EMBC.2012.6346525.

[13] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, “SMOTE: Synthetic Minority Over-sampling Technique,” Journal of Artificial Intelligence Research, vol. 16, no. 16, pp. 321–357, Jun. 2002, doi: 10.1613/jair.953.

 DOI 10.15203/99106-194-6-15

# Frequency-Resolved AC-Susceptibility Approach for MNPs Detection in Biomedical Applications

G. Barbieri*, L. Bangert, J. Arbustini, E. Elzenheimer, F. Wiekhorst, P. Radon, M. Gerken*

***Abstract*—Precise, non-invasive characterization of magnetic nanoparticles (MNPs) is key for biomedical applications such as targeted drug delivery and *in-vivo* protein monitoring. We present a portable and unshielded gradiometer setup capable of performing multi-frequency characterization of MNPs with the sample positioned outside the excitation and pick-up coils, employing an AC susceptibility approach. In addition to conventional in-coil characterization, the setup demonstrates reliable MNPs detection when a sample droplet is pumped and stopped near the sensing region at depths up to 10 mm. The proposed system allows a fast and wideband frequency sweep from 100 Hz to 10 kHz, keeping a low magnetic excitation field amplitude less than 0.5 mT and a total acquisition time of 8 minutes. The excitation and detection by a digital control architecture allows a customizable frequency scanning and a phase-sensitive signal demodulation. Validation was performed using micromod BNF-starch MNPs with a hydrodynamic diameter of 80 nm. The measured frequency-dependent response exhibited the expected dispersion behavior with a characteristic frequency of approximately 313 Hz. This work highlights the capability of the system to quantitatively characterize MNPs across a wide frequency range under realistic conditions at $d_{\mathrm{ch}} = 10\,\mathrm{mm}$ from the sensor, potentially suitable for *in-vivo*-oriented frequency-resolved analysis.**

***Index Terms*—ACS frequency sweep, gradiometer configuration, magnetic nanoparticles**

## I. Introduction

Magnetic nanoparticles (MNPs) have gained increasing attention for biomedical applications such as targeted drug delivery and *in-vivo* protein detection [1]. The importance of characterizing an MNP system under real and dynamic conditions is becoming significant, especially where the identification of this system allows specific bindings to be distinguished according to the customized target of the application [2].
Conventional techniques such as transmission electron microscopy (TEM) or magnetometry with superconducting quantum interference devices (SQUID) provide high-resolution information on particle structure and magnetism, but they are invasive, time-consuming, and not suitable for *in-vivo* or real-time measurements [3], [4]. Magnetic particle spectroscopy (MPS) systems, on the other hand, are excellent candidates for characterizing MNP systems, analyzing their structure by evaluating a large number of odd harmonics through the excitation of MNPs with relatively high magnetic fields [5], [6], [7]. Alternatively, AC susceptibility (ACS)-based commercial devices provide valuable information on the magnetic relaxation mechanisms of nanoparticles, potentially distinguishing Brownian and Néel contributions without requiring high magnetic fields or cryogenic operations and sweeping the frequency up to MHz range [8].
In this context, the development of portable, unshielded magnetic detection devices represents a promising alternative, allowing non-invasive and frequency-resolved characterization of MNPs in conditions closer to their biomedical application environment. The proposed system excites MNPs with magnetic fields lower than 0.5 mT and offers a time-efficient solution for the analysis of MNP systems suitable for *in-vivo* application where the Brownian dynamics is predominant with relatively slow relaxation times. As shown in Fig. 1, compared to typical ACS setup proposed in the literature [9], the presented frequency-resolved ACS system allows the characterization of a MNP sample outside the measurement box without shielding limits or in-coil constraints at a certain depth $d_{\mathrm{ch}}$. Here, $d_{\mathrm{ch}}$ denotes the distance from the sample volume center to the plane of the nearest pickup coil, measured along the sensitive axis. The MNP sample can be moved as a bolus orthogonally to the coils axis and characterized over a wide frequency range (100 Hz - 10 kHz). The configuration maintains full portability and compatibility with potential *in-vivo* conditions as presented in [10], [11], allowing for monitoring of MNP behavior in fluid without the need for sample preparation or magnetic shielding.

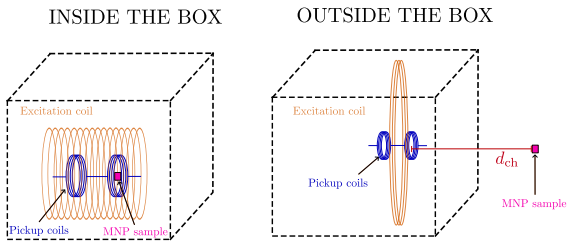


Fig. 1: The figure on the left shows a typical ACS setup where the sample is placed inside the sensing coil. On the right, the proposed setup is completely unshielded and involves placing the MNP sample away from the sensor at a distance $d_{\mathrm{ch}}$.

## II. Methods

### A. AC Susceptibility Approach

When a MNP system is exposed to an alternating magnetic field $H(t) = H_0 \sin(\omega t)$, the magnetization $M(t)$ follows

*Corresponding authors: G. Barbieri, M. Gerken (e-mail: giba, mge@tf.uni-kiel.de).
G. Barbieri, L. Bangert, J. Arbustini, E. Elzenheimer, M. Gerken are with the Kiel University, Germany
F. Wiekhorst, P. Radon, E. Elzenheimer (2nd-Afflination) are with the Physikalisch- Technische Bundesanstalt, Berlin, Germany

the excitation with a phase delay $\phi$, which depends on the effective relaxation process. The resulting magnetization can be expressed by the complex susceptibility $\chi(\omega) = \chi' - j\chi''$, where $\chi'$ and $\chi''$ are the in-phase and out-of-phase components and can be represented in the Debye theory [12] as follows:

$$\chi^{'}(\omega) = \frac{\chi_0}{1+(\omega\tau_{\text{eff}})^2}\,,\ \chi^{''}(\omega) = \chi_0\frac{\omega\tau_{\text{eff}}}{1+(\omega\tau_{\text{eff}})^2} \quad (1)$$

In (1), the static susceptibility is represented by $\chi_0$ and the effective relaxation time by $\tau_{\text{eff}} = \frac{\tau_{\text{B}}\tau_{\text{N}}}{\tau_{\text{B}}+\tau_{\text{N}}}$ which takes into account the combined effect of two distinct relaxation times: the Brownian ($\tau_{\text{B}}$) and Néel ($\tau_{\text{N}}$) relaxation times [12]. $\tau_{\text{B}}$ depends on the hydrodynamic volume and the medium viscosity, while $\tau_{\text{N}}$ depends on the anisotropy constant and the magnetic core volume.

A frequency sweep of the AC excitation field allows the frequency-dependent magnetic response of MNPs to be determined. $\chi''$ reaches its maximum near the characteristic relaxation frequency $f_{\text{ch}} = \frac{1}{2\pi\tau_{\text{eff}}}$, providing information about particle size and viscosity. In this work, $f_{\text{ch}}$ was estimated as the frequency at which the measured $\chi''(f)$ reaches its maximum (after the same interpolation/fitting procedure described in Sec. II-C).

### *B. Frequency-Resolved Technique Used with the Portable Gradiometer Setup*

The setup employs a first-order differential gradiometer, previously developed for MNP detection under dynamic flow conditions [11], composed of two symmetric pickup coils in antiphase, providing strong suppression of the homogeneous AC excitation field generated by a $25\,\text{cm}$ diameter coil.

For accurate frequency-resolved AC susceptibility measurements, the excitation magnetic field must remain constant across the entire frequency range. Any variation in the driving current amplitude would directly affect the field intensity inside the excitation coil and consequently distort the extracted susceptibility spectra. To guarantee a constant excitation field (maximum deviation of less than $1\,\%$), the coil current was monitored with a digital multimeter (34465A, Keysight technologies, Santa Rosa, USA) placed in series with the amplifier output, and the calibration factor was iteratively refined at each frequency to compensate for the frequency-dependent impedance of the coil. A 4-quadrant voltage and current amplifier (A1110-40 QE-100, Dr. Hubert GmbH, Bochum, Germany) was employed to drive the excitation coil with a sinusoidal current. In the present configuration, an optimized current of approximately $I_{\text{rms}} = 0.78\,\text{A}$ was applied, corresponding to a magnetic field amplitude of about $100\,\mu\text{T}$ at the coil center. A frequency vector for calibration purpose and for MNPs sample measurement routine was generated in logarithmic spacing from $100\,\text{Hz}$ to $10\,\text{kHz}$ with 30 discrete points.

A lock-in detection algorithm was implemented in MATLAB® (Version R2023b) to process the gradiometric signal acquired at each frequency of the sweep. The lock-in dection has been realized by the setup given in [11], with $f_{\text{s}} = 192\,\text{kHz}$; per frequency point we acquired $T = 10\,\text{s}$. A low-noise preamplifier (SR570, Stanford Research Systems, Sunnyvale, USA) was used to amplify the signal with a linear gain of 20, while simultaneously enabling programmable bandpass filters centered around the fundamental excitation frequency. The filter bandwidth was adjusted proportionally to the reference frequency to ensure optimal spectral preservation of the gradiometer signal without losing information on the fundamental component. This configuration effectively maximized the signal-to-noise ratio (SNR) during demodulation.

### *C. Samples and measurements*

A typical measurement procedure consisted of three sequential steps. First, a baseline acquisition was performed with an empty sample holder. Then, a reference paramagnetic sample ($Dy_2O_3$) with a known magnetic moment and zero imaginary susceptibility contribution was measured over the entire frequency vector for system calibration. Finally, the MNPs sample was characterized under identical conditions. The real and imaginary parts of the volumetric magnetic susceptibility of the MNP sample were extracted through a normalization process:

$$Z_{\text{norm}} = \alpha\left(\frac{Z_{\text{MNP}} - Z_{\text{empty}}}{Z_{\text{cal}} - Z_{\text{empty}}}\right),\ \alpha = \frac{m_{\text{cal}}\mu_0}{V_{\text{MNP}}B_{\text{exc}}} \quad (2)$$

In this expression, Z represents the measured complex voltage for the empty, calibration and MNP sample. The conversion factor $\alpha$ is proportional to the known magnetic moment of the calibration sample $m_{\text{cal}}$, the free space permeability $\mu_0$, the measured MNP volume $V_{\text{MNP}}$ and the exciting magnetic field amplitude $B_{\text{exc}}$.

For the frequency characterization, a water-diluted MNP system based on micromod BNF starch particles with a hydrodynamic diameter of $80\,\text{nm}$ was measured. Initially, a total sample volume of $50\,\mu\text{L}$ was placed inside a polycarbonate capsule and positioned within the sensing coil of the gradiometric sensor ($d_{\text{ch}} = 0\,\text{mm}$). Then, a sample volume of $50\,\mu\text{L}$ was pumped to the sensing region and positioned parallel to the sensing coil of the gradiometric detection system at a $d_{\text{ch}} = 10\,\text{mm}$.

In the post-processing stage, the susceptibility curves were fitted using spline interpolation with 500 points to increase the frequency resolution and enable the calculation of similarity parameters between curves. A quantitative comparison of the susceptibility curves measured with the MNPs sample at $d_{\text{ch}} = 0\,\text{mm}$ and $d_{\text{ch}} = 10\,\text{mm}$ was performed through similarity metrics, in particular evaluating the Root Mean Square Error (RMSE). RMSE was computed between the normalized curves over 100 Hz–10 kHz using the 500-point log-spaced spline-interpolated representation (same frequency grid for both positions).

## III. Results and Discussion

As shown in Fig. 2, the real and imaginary components of the measured complex susceptibility exhibit the expected frequency-dependent behavior. For the in-coil configuration ($d_{\text{ch}} = 0\,\text{mm}$), the characteristic relaxation frequency is approximately $313\,\text{Hz}$. When the sample is positioned at a

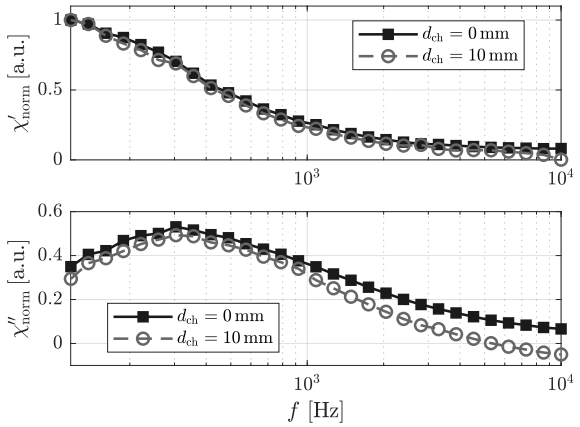


Fig. 2: ACS comparison results: the real and imaginary part of both susceptibility curves are normalized by the maximum of the real part signal and the two curves for $d_{\mathrm{ch}} = 0\,\mathrm{mm}$ and $d_{\mathrm{ch}} = 10\,\mathrm{mm}$ are shown.

distance $d_{\mathrm{ch}} = 10\,\mathrm{mm}$, the characteristic frequency exhibits a slight shift of around $8\,\mathrm{Hz}$ (estimated from spline-interpolated $\chi''$ peak location; uncertainty $\approx 2.6\,\mathrm{Hz}$), confirming that the setup maintains accurate frequency-resolved detection even under displaced measurement conditions.

The mean RMSE values indicate that the deviation between the two curves is more pronounced for the real part of the susceptibility ($\mathrm{RMSE}(Re[\chi]) = 0.07$); this effect can be attributed to phase calibration inaccuracies, which mainly affect the in-phase component of the signal. On the other hand, the RMSE of the imaginary part is clearly lower ($\mathrm{RMSE}(Im[\chi]) = 0.03$), confirming the higher similarity observed in the out-of-phase response. The characteristic frequencies calculated for the two $d_{\mathrm{ch}}$ differ by approximately $2.6\,\%$, which confirm the capability of the system to characterize MNP systems at a certain depth from the sensing coil. Nevertheless, the SNR plays a crucial role when the MNP sample is positioned far from the sensor, making the normalization increasingly sensitive at low frequencies (below $160\,\mathrm{Hz}$). Furthermore, considering 30 frequency points as in the presented measurements and taking into account the processing time between each frequency acquisition, the total acquisition time is approximately 8 minutes, allowing for an accurate evaluation of the characteristic frequency and the broadening of the imaginary component of the susceptibility curve. This remains well below the typical measurement times of conventional devices such as a Dynomag ACS device [13], [14], where an average acquisition time of 90 seconds per point would result in a total duration of approximately 45 minutes for the same number of frequency points used in this work.

## IV. Conclusion

These results show the capability of the developed gradiometric ACS configuration to reliably characterize MNPs systems over a wide frequency range with sufficient accuracy.

This work demonstrates that the proposed gradiometer-based setup is effectively employed for rapid (less than 10 seconds per frequency point) and quantitative frequency domain analysis of MNP samples outside the sensor system, providing a consistent assessment of the characteristic frequency and dispersive behavior of the magnetic response.

## Acknowledgment

This research was funded by the German Research Foundation (Deutsche Forschungsgemeinschaft) through project A8 and B12 of the Collaborative Research Center (CRC)1261 “Magnetoelectric Sensors: From Composite Materials to Biomagnetic Diagnostics”.

## References

[1] K. M. Krishnan, “Biomedical nanomagnetics: A spin through possibilities in imaging, diagnostics, and therapy,” *IEEE Transactions on Magnetics*, vol. 46, no. 7, pp. 2523–2558, 2010.

[2] S. E. Sandler, B. Fellows, and O. Thompson Mefford, “Best Practices for Characterization of Magnetic Nanoparticles for Biomedical Applications,” *Analytical Chemistry*, vol. 91, no. 22, pp. 14 159–14 169, 2019.

[3] K. R. Hurley, H. L. Ring, H. Kang, N. D. Klein, and C. L. Haynes, “Characterization of Magnetic Nanoparticles in Biological Matrices,” *Analytical Chemistry*, vol. 87, no. 23, pp. 11 611–11 619, 9 2015.

[4] A. Petri-Fink and H. Hofmann, “Superparamagnetic iron oxide nanoparticles (SPIONs): From synthesis to in vivo studies - A summary of the synthesis, characterization, in vitro, and in vivo investigations of SPIONs with particular focus on surface and colloidal properties,” *IEEE Transactions on Nanobioscience*, vol. 6, no. 4, pp. 289–297, 12 2007.

[5] T. Viereck, S. Draack, M. Schilling, and F. Ludwig, “Multi-spectral Magnetic Particle Spectroscopy for the investigation of particle mixtures,” *Journal of Magnetism and Magnetic Materials*, vol. 475, pp. 647–651, 4 2019.

[6] H. J. Krause and U. M. Engelmann, “Fundamentals and Applications of Dual-Frequency Magnetic Particle Spectroscopy: Review for Biomedicine and Materials Characterization,” 4 2025.

[7] S. B. Trisnanto, T. Kasajima, T. Shibuya, and Y. Takemura, “Low-Concentration Magnetic Particle Spectroscopy Using Gradiometric Receive Coil-Coupled Magnetoresistive Sensor,” *IEEE Transactions on Magnetics*, vol. 59, no. 11, pp. 1–5, 2023.

[8] E. Riordan, J. Blomgren, C. Jonasson, F. Ahrentorp, C. Johansson, D. Margineda, A. Elfassi, S. Michel, F. Dell’Ova, G. M. Klemencic, and S. R. Giblin, “Design and implementation of a low temperature, inductance based high frequency alternating current susceptometer,” *Review of Scientific Instruments*, vol. 90, no. 7, pp. 1–6, 2019.

[9] F. Ludwig, A. Guillaume, M. Schilling, N. Frickel, and A. M. Schmidt, “Determination of core and hydrodynamic size distributions of CoFe 2O4 nanoparticle suspensions using ac susceptibility measurements,” *Journal of Applied Physics*, vol. 108, no. 3, 2010.

[10] G. Barbieri, J. Arbustini, L. Ignacio, A. Acharán, A. Bahr, and M. Gerken, “Gradiometer-based Assessment of Magnetic Nanoparticles Quantification,” *International Conference on Electromagnetics in Advanced Applications (ICEAA)*, 2024.

[11] G. Barbieri, J. Arbustini, E. Elzenheimer, F. Wiekhorst, P. Radon, A. Teplyuk, R. Rieger, M. Höft, A. Bahr, and M. Gerken, “Unshielded Portable Gradiometer Setup for Magnetic Nanoparticles Characterization in Flow Condition,” 8 2025. [Online]. Available: https://www.techrxiv.org/users/949926/articles/1320075-unshielded-portable-gradiometer-setup-for-magnetic-nanoparticles-characterization-in-flow-condition?commit=b04427e54dab055cb4c7e7a195e6604007c86528

[12] N. J. Darton, A. Ionescu, and J. Llandro, *Magnetic Nanoparticles in Biosensing and Medicine*. Cambridge University Press, 2 2019.

[13] S. Liu, A. Heshmat, J. Andrew, I. Barreto, and C. M. Rinaldi-Ramos, “Dual imaging agent for magnetic particle imaging and computed tomography,” *Nanoscale Advances*, vol. 5, no. 11, pp. 3018–3032, 5 2023.

[14] S. Sepehri, J. Andersson, V. Schaller, C. Grüttner, M. Stading, and C. Johansson, “Remote Sensing of the Nano-Rheological Properties of Soft Materials Using Magnetic Nanoparticles and Magnetic AC Susceptometry,” *Nanomaterials*, vol. 13, no. 1, 1 2023.

 DOI 10.15203/99106-194-6-16

# MPI Phantoms for Inter-Scanner Comparison

M. Löffler, B. Kluwe, C. Fiedler, F. Wiekhorst, S. Dutz

***Abstract*— Measurement phantoms are well-established tools in many areas of medical diagnostics. However, in the field of magnetic particle imaging, there is a lack of standardized phantom structures and operating protocols. In this work, we describe the development of phantoms aiming at closing this gap. For this purpose, we designed static phantoms that can be filled with long-term stable materials, as well as flow phantoms controlled by a microfluidic control system.**

***Index Terms*—measurement phantoms, magnetic particle imaging, 3D-printing**

## I. Introduction

Measurement phantoms are well-established tools to analyze medical imaging modalities [1]. The performance of devices, algorithms, and new methods could be explored and compared by using them. Well-established phantoms are available for imaging methods in medical diagnostics, like computer tomography and magnetic resonance imaging. A new technique called magnetic particle imaging (MPI) is an emerging imaging modality with promising potential in many fields of medical diagnostics [2]. Since its first demonstration in 2005, various hardware architectures, parameter configurations, reconstruction methods, and use cases have been developed, tested, and explored. To evaluate new devices and methods, different kinds of phantoms are used [3]. The phantoms often work well for the given situation, but a comparison of the results is difficult due to the lack of defined phantom structures and standardized quantification protocols [4]. To close these gaps, we are developing phantoms in an iterative process to use them within different MPI scanners. The results will be utilized to identify phantoms that are usable in different scanner architectures for comparison and to develop uniform working and calibration procedures. This will be included in a planning tool to find a suitable scanner for a given medical use case.

## II. Methods

In the first iteration, we developed phantoms based on data found in literature of the MPI research community. A stackable static phantom (SSP) and flow phantoms were developed. The SSP consists of stackable plates that contain commonly used structures in published phantoms, like letters, line and point sources and different shapes that are fillable with long-term stable structure. Two alternative ways of filling the molds are envisaged. Method one is a silicon-matrix filled with magnetic nanoparticles (MNP).

For the second method we directly include MNPs during a 3D-printing process to obtain 3D-printed geometry with defined properties for use in MPI. The geometries of all printed structures were scanned with the Revopoint MINI 3D-Scanner to analyze their geometry. The structures containing nanoparticles can thus be compared with the MPI images and the constructed model. For the flow phantom, we developed three different structures. Two phantoms consist of two pipes with different geometries (see Fig. 1c and d) and the third based on a segmented vessel structure (see Fig. 1 e). The MNPs containing dispersions are set through micromixing and pumped through the phantoms controlled by a microfluidic control system, which enables dynamic changes in MNP dispersion concentration.

## III. Results

Figure 1a shows the SSP, Figure 1b the whole setup of the flow phantom, and in Figure 1c-e the three different flow phantoms are shown. The SSP could be extended by adding plates with individual defined structures. The molds can be filled with long-term stable silicone or 3D-printed structures like described in [5] and [6]. The developed flow phantoms allow the measurement of spatial and temporal resolution, as well as sensitivity. Therefore, the concentration of MNPs could be adjusted using the microfluidic control system. Flow phantom one consists of two parallel pipes with an axial difference of 2.5 mm. Flow phantom two is a V-shaped phantom. By including an MNP-containing reference structure with defined spaces between the reference object and the axes,

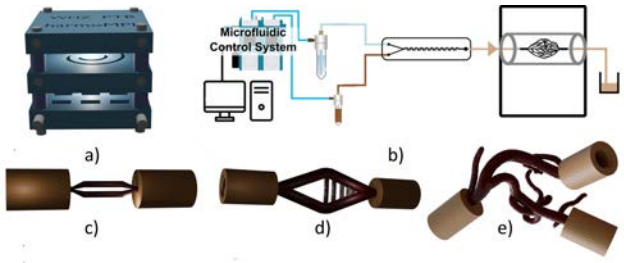


Fig. 1. a) Expandable, stackable phantom (SSP) with three plates. The lower plate includes 4 mm × 2 mm × 2 mm molds with spacing from 1 mm to 2.5 mm in 0.5 mm increments. b) Schematic setup of the microfluid control system. c) Two-pipe phantom with two parallel pipes whose center-to-center distance is 2.5 mm. d) Flow phantom in a V-shape that enables parallel assessment of spatial and temporal resolution as well as sensitivity. e) MRI-based vascular segmentation flow phantom.

M. Löffler, C. Fiedler, S. Dutz, Faculty of Physical Engineering/Computer Sciences, University of Applied Science Zwickau, Germany,
e-mail: marcus.loeffler@whz.de

B. Kluwe and F. Wiekhorst, Physikalisch-Technische-Bundesanstalt, Berlin, Germany

S. Dutz, Institute of Biomedical Engineering and Informatics, TU Ilmenau, Ilmenau, Germany

we can measure the spatial resolution by finding the minimum space that can be resolved by a given MPI system. Figure 2a shows MPI images of a helix structure located in different surroundings. A comparison between an STL file of a 3D-printed helix structure and a scan (Figure 2b) is shown in Figures 2c and 2d. The surface from the STL file was resampled using Poisson-disc sampling with 10k points. The resulting point cloud was then compared with the point cloud obtained from the scan.

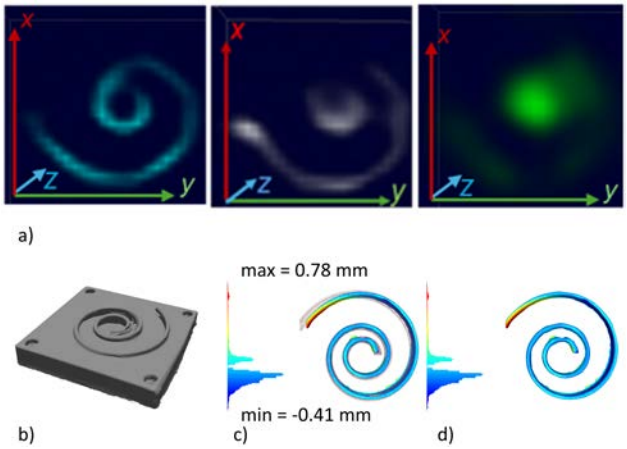


Fig. 2. a) MPI images of the 3D-printed helix structure acquired under three conditions: in fluid environment, after immobilization by freeze-drying, and following cellular uptake. b) Scanned helical structure positioned on a plate. c) Difference image obtained by comparison between the scanned helix structure and the STL file. The gridded helix represents the originally constructed geometry from the STL file. d) Difference image without grid. The maximum difference is 0.78 mm and the minimum is -0.41 mm in c and d.

## IV. Conclusion

In this work, a stackable phantom and three flow phantoms were developed and presented. The phantoms are fillable with long-term stable material containing MNPs. A geometry-scanning method to be used in future work to compare printed geometries with the MPI images is presented. Future MPI measurements will be conducted, and the phantoms will be further refined during the research project “harmoMPI”.

## Acknowledgment

This work was founded by the DFG project “harmoMPI” (project no 540326102)

## References

[1] L. B. Christie *u. a.*, „Review of imaging test phantoms“, *J. Biomed. Opt.*, Bd. 28, Nr. 08, Aug. 2023, doi: 10.1117/1.JBO.28.8.080903.

[2] A. Davida und B. Basari, „The hardware, tracer, and signal processing methods of magnetic particle imaging: A review“, *J. Appl. Phys.*, Bd. 136, Nr. 22, S. 220701, Dez. 2024, doi: 10.1063/5.0220219.

[3] E. U. Saritas *u. a.*, „Magnetic Particle Imaging (MPI) for NMR and MRI researchers“, *J. Magn. Reson.*, Bd. 229, S. 116–126, Apr. 2013, doi: 10.1016/j.jmr.2012.11.029.

[4] E. Ureña Horno *u. a.*, „Strategies towards standardizing calibration methods for magnetic particle imaging (MPI) signal quantification: solution *vs.* cellular environments“, *Nanoscale*, Bd. 17, Nr. 41, S. 24060–24071, 2025, doi: 10.1039/D5NR03025K.

[5] L. Wöckel *u. a.*, „Long-term stable measurement phantoms for magnetic particle imaging“, *J. Magn. Magn. Mater.*, Bd. 471, S. 1–7, Feb. 2019, doi: 10.1016/j.jmmm.2018.09.012.

[6] N. Löwa, J.-M. Fabert, D. Gutkelch, H. Paysen, O. Kosch, und F. Wiekhorst, „3D-printing of novel magnetic composites based on magnetic nanoparticles and photopolymers“, *J. Magn. Magn. Mater.*, Bd. 469, S. 456–460, Jan. 2019, doi: 10.1016/j.jmmm.2018.08.073.

 DOI 10.15203/99106-194-6-17

# FPGA IQ Demodulation and GNU Radio Acquisition for Converse ME Sensors Using the RedPitaya SDR Platform

E. Miranda†, J. Arbustini†, P. Mendoza, E. Elzenheimer, H. Wolframm, M. Höft, R. Rieger, A. Bahr

***Abstract*—Amplitude-modulated (AM) magnetic field sensors, such as the converse Magnetoelectric (converse ME, cME), detect extremely weak magnetic field signals by amplitude-modulating a low-frequency signal onto sidebands around a high-frequency (hundreds of kHz) carrier that occurs near its mechanical resonance. However, digital readout of these sensors remains limited by the lack of open, continuous-stream demodulation systems capable of handling the ultra-low modulation indices ($m \leq 0.01$) required for biomedical applications such as magnetocardiography (MCG). This work presents a continuous data-acquisition IQ demodulation system based on the RedPitaya SDRlab 122-16 platform, using the AMD Zynq-7020 System-on-Chip (SoC) and the GNU Radio environment. The Hardware-Description-Language (HDL) designed in System Verilog performs coherent IQ demodulation and decimation in the FPGA (Artix-7), and the AMD processor (dual-core Arm A9-Cortex) streams the data over an Ethernet connection, enabling continuous monitoring. The system is characterized using single-tone AM stimuli over an $m = 10^{-1}$, with linear reconstruction (after demodulation in the baseband) demonstrated down to $m = 10^{-4}$ for the emulator setup, and $m \geq 10^{-2}$ for the cME sensor setup, both limited by their noise-floor. Finally, a proof-of-concept measurement using a cME sensor validates the magnetic-sensing solution. This approach offers a low-cost, open-source, and reproducible pathway for multichannel ME readout and real-time monitoring of (bio-)magnetic signals.**

***Index Terms*—Amplitude modulation, biomagnetic sensing, system-on-chip, magnetometers, digital design, sensor readout.**

## I. Introduction

Modern biomagnetic sensing relies on digital acquisition systems that bridge analog physiological phenomena with advanced signal processing algorithms. Magnetoelectric (ME) sensors, particularly Amplitude-Modulated (AM) ME (converse ME, cME) magnetic field sensors, provide a compelling path toward non-invasive biomagnetic diagnostics, offering non-cryogenic operation and the ability to detect fields in the picoTesla (pT) range [1]. Unlike Electrocardiography (ECG), Magnetocardiography (MCG) can detect the cardiac vector field without skin contact, electrode preparation, or impedance variability, underscoring the need for compact, low-noise magnetometer readout systems.

The cME sensors operate by exciting a thin-film multilayer cantilever at its mechanical resonance (four U-mode resonance frequencies in the range from $400$ to $600\,\mathrm{kHz}$), so that low-frequency magnetic fields modulate the oscillation amplitude, and readout inductively using a pick-up coil. The resulting AM (a double-sideband AM with carrier) signal contains extremely weak sidebands, often in the range of $40$ to $60\,\mathrm{dB}$ below the carrier. The equivalent modulation indices $m$ are in the order of $0.001$ to $0.01$. Recovering these sidebands requires demodulators with high dynamic range, minimal phase error, and the ability to operate continuously without data gaps at sampling frequencies of megasamples per second ($\mathrm{MSa/s}$).

Conventional 14 to 16-bit solutions, such as lock-in amplifiers (e.g., Zurich Instruments products, Zurich, Switzerland), offer excellent demodulation but are expensive and limited in scalability across multiple channels. Other commercial Software-Defined Radio (SDR) or Data Acquisition (DAQ) platforms with $\geq 100\,\mathrm{MSa/s}$ conversion speeds rarely provide an open FPGA fabric for custom demodulation (e.g., Spectrum Instrumentation GmbH, Grosshansdorf, Germany). Prior work using a Field-Programmable Gate Array (FPGA) demonstrated the feasibility of In-phase and Quadrature (IQ) demodulation but suffered from intermittent sampling, gaps in Block Random Access Memory (BRAM) or Double Data Rate Synchronous Dynamic RAM (DDR) buffering, and insufficient throughput for continuous $\mathrm{MSa/s}$ acquisition directly through the computer host [2, 3]. Hardware-in-the-loop (HIL) approaches in Simulink (Mathworks, Massachusetts, United States), such as FPGA-in-the-loop (FIL) or Processor-in-the-loop (PIL), lack design files for custom boards, such as the Redpitaya SDRlab 122-16. The open-source environment from Redpitaya offers an instrumentation device. Still, it is limited to long-term continuous data streaming (as required in some applications) and to lower-frequency resolution for frequency-domain analysis ($< 1\,\mathrm{Hz}$). Overall, this highlights the gap of a continuous-stream FPGA solution for AM demodulation of cME sensors that supports $\mathrm{MSa/s}$. Therefore, in this work, an IQ demodulation system with decimation is proposed to provide stable long-term streaming without BRAM underflow/overflow, and it integrates with *GNU Radio*, a free and open-source signal processing runtime and signal processing software development toolkit.

## II. Methods

In this work, the FPGA performs real-time demodulation with a fully pipelined datapath clocked at the analog-to-digital converter (ADC) sampling rate ($f_\mathrm{s} = 122.88\,\mathrm{MHz}$; $T_\mathrm{s} \approx 8.14\,\mathrm{ns}$), sustaining continuous sample throughput without stalls or dropped samples. The processing latency is fixed and deterministic.

†E. Miranda and J. Arbustini are co-first authors. Corresponding author: J. Arbustini (e-mail: jrsa@tf.uni-kiel.de)

J. Arbustini, E. Elzenheimer, H. Wolframm, M. Höft, R. Rieger are with the Kiel University, Kiel Germany.

E. Miranda and P. Mendoza are with the Instituto Tecnològico de Costa Rica, Cartago, Costa Rica

A. Bahr is with the Technische Universität Dresden, Germany.

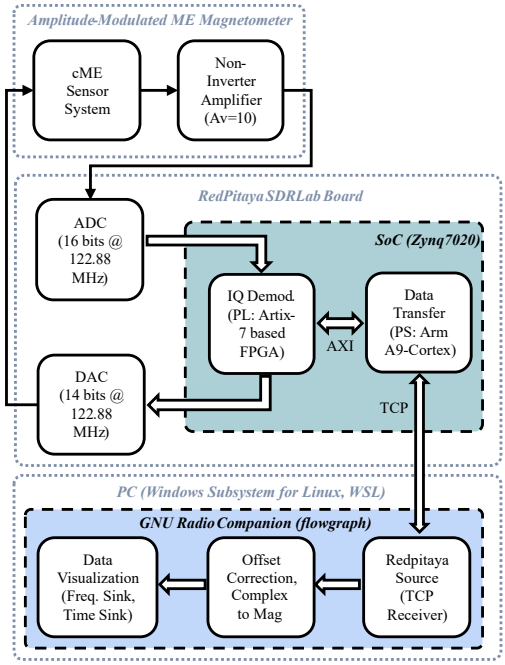


Fig. 1: Simplified diagrams of the FPGA IQ Demodulation (Hardware) and GNU Radio Acquisition (Software) system for the cME Sensor using the SoC Zynq7020 in the RedPitaya SDRlab 122-16 platform.

### A. System Architecture

The architecture overview is shown in Fig. 1. The system leverages the RedPitaya SDRlab 122-16 (RedPitaya, Solkan, Slovenia), based on a Zynq-7020 System-on-Chip (AMD Zynq™ 7000 SoC, Santa Clara, United States). Its Programmable Logic (PL) contains the Artix-7 based FPGA, and its Processing System (PS) has a dual-core Arm A9-Cortex running Linux. The ADC within the Redpitaya has a resolution of $16\,\mathrm{bits}$, while the Digital-to-Analog Converter (DAC) resolution is $14\,\mathrm{bits}$. Both, the ADC and DAC run at a sampling rate of $122.88\,\mathrm{MSa/s}$. The DAC provides a signal to excite the cME sensor at its resonance frequency and with the required voltage amplitude. Unlike fixed-function ADC front-ends, this platform enables custom Hardware Description Language (HDL) for continuous IQ demodulation and decimation (down-conversion of the sampling rate) [3]. The processing system uses the Advanced eXtensible Interface (AXI) on-chip communication bus protocol and Transmission Control Protocol (TCP) for data transfer. With TCP-based transfer and implemented buffering, we observed continuous streaming without sample gaps in our experiments [2]. At the same time, Windows Subsystem for Linux (WSL) provides seamless integration with GNU Radio and network tools without requiring a full operating system (OS) transition, and GNU Radio handles higher-level digital signal-processing algorithms for visualization and storage. This digital amplitude demodulator readout design is based on the open-source SDR transceiver framework, and extended to provide a specialized GNU Radio flowgraph for this application [4].

The FPGA IQ Demodulator-based system is explained in detail in [3, 4]. This implementation avoids the limitations on amplitude and phase dependency of analog approaches, where analog mixers focus on frequency conversion to an intermediate frequency or baseband before digitization [5]. Nevertheless, the drawback is that high-sampling-rate ADCs offer fewer bits due to design trade-offs in integrated circuits. The digital design modules are:

**Numerically Controlled Oscillator (NCO):** A Direct Digital Synthesis (DDS)-based NCO generates sin/cos waveforms at the cME sensor mechanical resonance (in the range of $400$ to $600\,\mathrm{kHz}$). In this case, a $30\,\mathrm{bits}$ phase accumulator yields sub-Hz tuning resolution, essential for matching the exact resonant frequency of each cME cantilever [6].

**Coherent Mixing and Low-Pass Filtering:** The ADC stream is mixed by the NCO to produce $I(t) = x \cdot \cos(2\pi f_{\mathrm{res,me}})$ and $Q(t) = x \cdot \sin(2\pi f_{\mathrm{res,me}})$. A cascaded CIC and FIR decimation filter chain reduces the sampling rate from $122.88\,\mathrm{MSa/s}$ to selectable $24$, $48$, $96$, $192$, $384$, $768$ or $1536\,\mathrm{kSa/s}$, preserving the envelope signal of interest [4]. In the FPGA-based IQ demodulator, the Verilog implementation produces $32\,\mathrm{bits}$ outputs for both the I and Q channels. The signals are subsequently decimated to kSa/s compatible with the TCP-based continuous data streaming.

**Continuous Streaming:** To avoid the BRAM overflow issues described in earlier work, data are transferred directly from the PL BRAM to the PS DDR via AXI, then streamed over TCP in fixed $16\,\mathrm{kB}$ frames per samples of $32\,\mathrm{bits}$ [2]. This method provides uninterrupted operation for measurement windows required to achieve frequency resolution of $\leq 1\,\mathrm{Hz}$.

**GNU Radio Flowgraph:** In addition, the acquisition chain includes a dedicated GNU Radio flowgraph tailored for this project. The GNU Radio application receives IQ packets, reconstructs the complex-valued stream, computes the envelope magnitude $\sqrt{I^2 + Q^2}$, and provides real-time visualization and traceable recordings for later analysis with specialized scripts.

### B. Experimental Characterization

Two classes of experiments were performed to validate the IQ demodulator: **Setup A — Controlled AM stimuli:** An emulator which mimics the cME sensor output signal (hardware setup using a vector signal generator R&S SMBV100A (Rohde & Schwarz, München, Germany) and a low-noise voltage source LS-M81-SSM-2 including LS-M81-VS-10 (Lake Shore Cryotronics, Inc., Westerville, OH, USA), described in detail in [5]) generates an AM signal with carrier $f_{\mathrm{c}} = 500\,\mathrm{kHz}$, and envelope signal $f_{\mathrm{b}}$ at $19.44\,\mathrm{Hz}$. In this setup, the modulation index $m$ can be varied within the high-dynamic-range of $120\,\mathrm{dB}$. The generator output was connected directly to the RedPitaya ADC input to measure the acquisition chain baseline. The analysis was performed using Welch's method with a segment length of 4096 samples, 50% overlap, and an FFT size of 4096, resulting in five averaging windows. **Setup B — cME sensor (Proof-of-Concept):** A prototype cME sensor from Kiel University (Kiel, Germany) in

 DOI 10.15203/99106-194-6-17

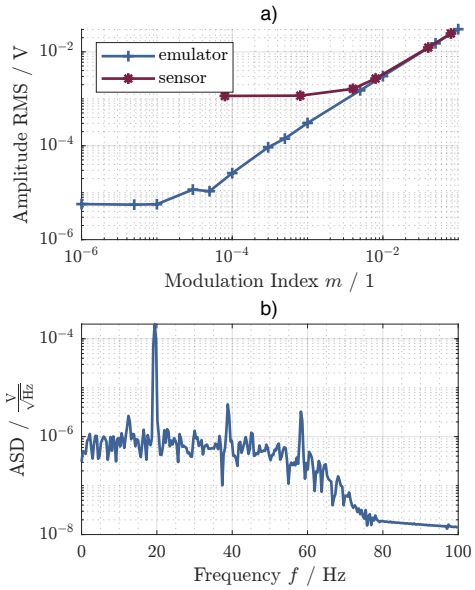


Fig. 2: Performance characterization of the real-time IQ demodulation system. (a) Recovered envelope RMS amplitude versus modulation index $m$ for both emulator and cME sensor; (b) Amplitude Spectral Density (ASD) using the mimicked AM signal of $m = 0.05\,\%$ and $f_\mathrm{b} = 19.44\,\mathrm{Hz}$.

a mobile magnetically-shielded cylinder, providing a shielding factor of 20 dB (Aaronia Zero Gauss-Chamber, Strickscheid, Germany). The cME was electrically excited near its resonance ($f_\mathrm{res} = 508\,\mathrm{kHz}$). The magnetic signal was generated with a solenoid coil (placed inside the shielded box) at $19.44\,\mathrm{Hz}$ excited with different amplitudes using a current generator Keithley 6221 (Keithley, Ohio, United States). The RedPitaya generates the excitation and receives the cME sensor output signal after the non-inverter amplifier, with a gain of $10\,\mathrm{V/V}$. The GNU Radio flowchart for both setups generates a data-streaming of $32\,\mathrm{bits}$ at $24\,\mathrm{kSa/s}$, with a Low-Pass-Filter (LPF) cutoff frequency at $60\,\mathrm{Hz}$.

## III. Results

Fig. 2(a) shows the recovered test signal after demodulation in the baseband using one-second measurement to compute the RMS amplitude versus $m$ for sweeps performed at $f_\mathrm{b} = 19.44\,\mathrm{Hz}$. For the vector generator, linear response was observed across $m = 10^{-4}$ to $10^{-1}$, with a limit-of-detection $\mathrm{LOD} = 7\,\mu\mathrm{V}$ calculate with the resulting noise variance as presented in [7]. The cME sensor measurements (carrier at $508\,\mathrm{kHz}$) showed comparable linearity above $m \sim 10^{-2}$, with $\mathrm{LOD} = 1.244\,\mathrm{mV}$, which reflects the additional overall noise present in the experimental setup using the cME sensor system (as suggested by previous studies [7]). Fig. 2(b) shows the Amplitude-Spectral-Density (ASD) of a AM of low-$m$ stimulus generated with the vector generator ($m = 0.05\%$, $f_\mathrm{b} = 19.44\,\mathrm{Hz}$). The spectrum is dominated by a single, well-defined peak at $f_\mathrm{b}$; remaining spectral lines appear only as low-amplitude harmonics above the noise floor. From this experimental measurements in Fig. 2(b), the total harmonic distortion (THD) is $-31.094\,\mathrm{dB}$. This indicates that the demodulation chain recovers the expected spectral content of the envelope for very small $m$, preserving the fundamental tone without introducing measurable spurious components from the carrier.

## IV. Discussion

The results demonstrate that the low-cost FPGA-based IQ demodulation, combined with GNU Radio, provides a method for recovering ultra-low-$m$ AM signals in cME sensing. From the comparison between the mimicked signal and the real sensors, the AM sideband amplitude (envelope) demodulated scales approximately with $20\,\mathrm{dB/dec}$ of $m$ in the linear region, consistent with theoretical scaling. In addition, it is shown that the detection threshold is primarily limited by overall cME sensor noise and coupling losses, rather than the proposed acquisition system. The measured THD (Total Harmonic Distortion) corresponds to a distortion level of approximately 2 - 3 % of the signal of interest, which remains within a tolerable range for cardiac event detection. For (bio-)medical applications, robust heartbeat timing can be achieved through event-detection algorithms, while the observed LoD reduction factor of 177 significantly improves the baseband SNR, which is more critical than THD for reliable MCG envelope reconstruction in the current stage of development.

## V. Conclusion

We developed a complete RedPitaya-based SDR acquisition system that performs coherent IQ demodulation and decimation in the FPGA for different $m$ of AM signals. In addition, the SoC allows the data streaming for the reconstruction of the signal after baseband conversion, and demonstrates compatibility for cME biomagnetic measurements. Thus, this solution supports continuous, high-throughput acquisition, minimizes data volume. Furthermore, the proposed system will allow to accurately characterize and enable the next generation of cME sensors with improved sensitivity. Future work would focus on resource and functional optimization towards integration of the system on an ASIC.

## Acknowledgment

This work was funded by the DFG through project B1 of CRC 1261 *"Biomagnetic Sensing"* (Grant 286471992).

## References

[1] E. Elzenheimer, P. Hayes, *et al.*, "Investigation of Converse Magneto-electric Thin-Film Sensors for Magnetocardiography," *IEEE Sensors J.*, 2023.

[2] A. Morales-Zamora, J. Arbustini, *et al.*, "Lossless 1.6 Mbits/s FPGA-based Real-Time Data Transmission System for Converse ME Sensors," in *IEEE CONCAPAN*, 2024.

[3] C. McGuinness-Rodriguez, J. Arbustini, *et al.*, "Resource-Efficient FPGA-based I/Q-Demodulator Detects 0.0005 Modulation Index with 9.5 dB SNR for Converse ME Sensors," in *IEEE CONCAPAN*, 2024.

[4] P. Demin, "SDR transceiver 12288." https://pavel-demin.github.io/red-pitaya-notes/sdr-transceiver-122-88/, 2021. Accessed: 2025-10-19.

[5] J. Arbustini, E. Elzenheimer, H. Wolframm, *et al.*, "Analog Demodulation for Electrically Modulated ME Sensors: A Signal Integrity Analysis and Evaluation Approach," *IEEE Sensors J.*, 2025.

[6] J. Munoz, J. Arbustini, *et al.*, "Digital Approaches on Frequency Tuning for Magnetoelectric Sensors," in *IEEE ICECS*, 2022.

[7] E. Elzenheimer *et al.*, "Quantitative evaluation for magnetoelectric sensor systems in biomagnetic diagnostics," *Sensors*, 2022.

 DOI 10.15203/99106-194-6-18

# Model-Based Design of Continuous-Time $\Sigma\Delta$ Modulators for Integrated Readout of Amplitude-Modulated Magnetometers

J. Arbustini, N. Matarrita, P. Mendoza, E. Elzenheimer, E. Böhme, M. Höft, R. Rieger, A. Bahr

***Abstract*—This work investigates the suitability of continuous-time sigma-delta ($\Sigma\Delta$) modulators for envelope reconstruction in Amplitude-Modulated (AM) magnetometers utilized in (bio-)magnetic sensing. Electrically excited Magnetoelectric (ME) sensors generate a weak AM signal with an ultra-low modulation index $m$, making the extraction of the envelope, which carries, for example, cardiac information for Magnetocardiography (MCG), particularly challenging. Accurate envelope recovery places stringent demands on the modulator topology, the attainable noise-shaping performance, and the linearity of the overall magnetometer readout at low magnetic flux densities. We assessed three third-order $\Sigma\Delta$ loop architectures (CIFB, CIFFB, and CIFF) using a unified model-based design flow implemented in MATLAB/Simulink. A comparative analysis using simulated sinusoidal signals and synthetic MCG waveforms as test inputs reveals the influence of the loop architecture on envelope fidelity, in-band noise performance, and the achievable linear operating range. The results indicate that all architectures can reconstruct the envelope for modulation indices down to $m$ = 0.001. CIFB achieves the lowest RMS noise floor at ultra-low $m$, while CIFFB provides the best envelope fidelity for the synthetic MCG signals. These results support CT $\Sigma\Delta$ modulators as a feasible integrated readout approach for amplitude-modulated magnetometers.**

***Index Terms*—amplitude-modulation, magnetometers magnetocardiography, (bio-)magnetic sensing, magnetoelectric sensors, sigma-delta modulation**

## I. Introduction

The development of electronic readout schemes for magnetometers used in non-invasive magnetic biosignal detection is particularly challenging, as both the sensing device (detector) and its associated electronics must resolve magnetic fields with exceptionally small magnetic flux densities. Magnetocardiography (MCG) is a prominent application where portable and scalable systems require an overall system behavior that offers high-sensitivity, low-noise spectral density, bandwidth, timing parameters, and compatibility for multichannel integration [1]. While Superconducting Quantum Interference Devices (SQUIDs) continue to serve as the benchmark (gold standard) for (bio-)magnetic measurements, their reliance on bulky Dewar-based cryogenic cooling restricts sensor flexibility, increases the source-to-detector distance, and limits portability. Non-cryogenic alternatives that operate without additional heating, including Tunneling Magnetoresistance (TMR) devices [2], Helium gas Optically Pumped Magnetometers ($^4$He OPMs) [3] and electrically modulated Magnetoelectric (ME) Sensors (converse ME, cME) [4], provide a promising route toward compact and scalable systems. Owing to their operating principle, cME sensors modulate the (bio-)magnetic signal, such as the MCG signal, onto the envelope of a kHz-range amplitude-modulated (AM) carrier. This upconversion improves immunity to low-frequency noise (1/$f$-noise) and disturbances, but imposes stringent requirements on the Analog-to-Digital Converter (ADC). Recovering microvolt-level envelope variations intrinsically demands low-noise operation, sufficiently high sampling rates to ensure accurate signal reconstruction, and robust suppression of carrier-related artifacts.

Continuous-Time (CT) sigma-delta ($\Sigma\Delta$) modulators are particularly well suited for the integration of cME magnetometer readouts due to their strong noise-shaping capability, relaxed analog filtering requirements, and compatibility with low-power CMOS technologies [5–8]. However, the role of selection of $\Sigma\Delta$ loop topology in determining baseband signal reconstruction performance at extremely low modulation indices $m$ encountered in AM-modulated magnetometers remains inadequately studied. In particular, systematic comparisons of Cascade-of-Integrators with FeedBack (CIFB), hybrid Cascade-of-Integrators in FeedForward-FeedBack (CIFFB), and Cascade-of-Integrators with FeedForward (CIFF) architectures, especially with respect to noise performance, and robustness for AM modulated (bio-)magnetic signals, have not yet been reported for AM-modulated cME magnetometer readouts at the ultra-low modulation indices $m$ considered here. This gap is addressed here by comparing three third-order CT $\Sigma\Delta$ modulator topologies using a simulated monofrequency signal and a synthetic MCG. This test signal represents the envelope signal after the AM sensor model, and therefore compares against the reconstruction at the baseband after CT $\Sigma\Delta$ modulator and the signal processing. The test program analyzes envelope reconstruction, noise-shaping behavior, and performance under idealized readout conditions within a unified model-based design framework implemented in MATLAB/Simulink (MathWorks, Inc., Massachusetts, United States). The workflow is further contextualized through a brief review of established coefficient-synthesis tools. Practical guidelines for selecting $\Sigma\Delta$ architectures suitable for future integrated, low-power ME magnetometer front ends are provided based on the results.

## II. Methods

In this work, MCG signals are modeled as low-frequency variations that appear in the *envelope* of an AM-modulated magnetometer output. The cME sensor produces a double-sideband AM signal with carrier $a_\mathrm{c}\sin(2\pi f_\mathrm{c}t)$, where the

Corresponding author: J. Arbustini (e-mail: jrsa@tf.uni-kiel.de)

J. Arbustini, E. Elzenheimer, M. Höft, R. Rieger are with the Kiel University, Kiel Germany.

N. Matarrita and P. Mendoza are with the Instituto Tecnològico de Costa Rica, Cartago, Costa Rica

E. Böhme and A. Bahr are with the Technische Universität Dresden, Germany.

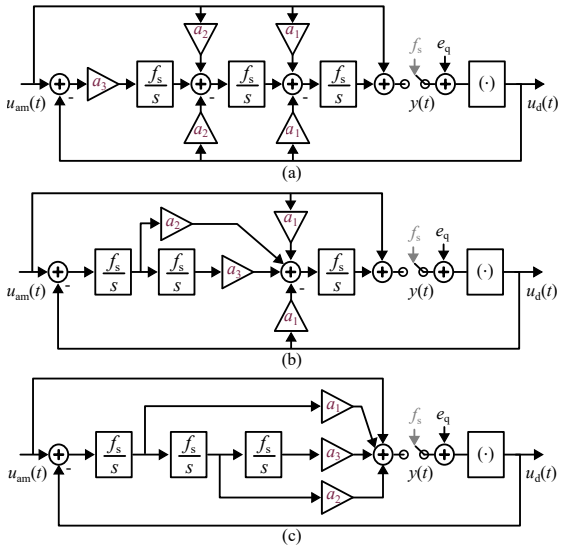


Fig. 1: Block diagrams of the evaluated third-order CT ΣΔ modulators: (a) CIFB, (b) CIFFB, and (c) CIFF.

envelope corresponds to the desired magnetic signal $b_\mathrm{d}(t)$. MCG amplitudes are extremely small, therefore the associated $m \ll 0.01$, requires accurate and low-noise demodulation to preserve diagnostic waveform morphology [9].

### A. Continuous-Time ΣΔ Modulator

The CT ΣΔ modulators suppress in-band quantization noise through oversampling and feedback. Under the standard linear model, the quantizer is represented as a gain plus additive white noise (AWGN), yielding

$$Y(z) = \mathrm{STF}(z)\,X(z) + \mathrm{NTF}(z)\,E(z),$$

where the Signal-Transfer-Function (STF) controls signal fidelity and the Noise-Transfer-Function (NTF) controls noise shaping. $X(z)$ denotes the $z$-transform of the input signal, and $E(z)$ represents the the $z$-transform of additive quantization noise (modeled as AWGN). For envelope extraction, a flat and near-unity STF is essential [5]. Fig. 1 shows the three CT third-order loop structures evaluated in this work. CIFB applies DAC feedback to every integrator, improving robustness; CIFFB combines CIFB and CIFF; and CIFF introduces feedforward paths, reducing internal swing. These architectural differences affect in-band noise, internal state swing, maximum stable amplitude (MSA), and overall stability [5]. The oversampling ratio (OSR) is defined relative to the carrier frequency as $\mathrm{OSR} = f_\mathrm{s}/(2 \cdot f_\mathrm{c})$, where $f_\mathrm{c}$ is defined as the carrier frequency equal to the supported ΣΔ bandwidth of $100\,\mathrm{kHz}$, assuming an intermediate-frequency conversion after the cME sensor readout. The multi-bit quantizer was modeled as a uniform midrise quantizer (15 output levels, symmetric about zero) with the boundary points in Simulink.

### B. Coefficient Synthesis Tools for ΣΔ Modulators

The typical loop coefficients determine the NTF, STF, and internal swing behavior, making their synthesis critical, particularly for future integrated circuit (IC) implementations where headroom, jitter, and integrator linearity impose strict constraints. In this work, coefficients were generated using an in-house MATLAB script based on the *ABCD state-space formulation*. This approach maps a target NTF to a realizable loop filter and performs state scaling to maximize MSA and avoid saturation. The workflow aligns with the methodologies implemented in Schreier's ΣΔ Toolbox [5–7] This tool analyzes NTF design, estimates stability, and automates state-space realization, making it ideal for early architecture exploration. Other established tools complement this synthesis flow such as: **SIMSIDES:** A Simulink-based environment supporting non-idealities (e.g. jitter, finite gain, excess-loop delay, DEM) and optimization algorithms, suitable for IC-level refinement [7]. **Ulm CT ΣΔ Synthesis Tool:** Uses lifting-based modeling with GPU/FPGA acceleration for rapid exploration of large CT ΣΔ design spaces [10]. These tools illustrate a design ecosystem that ranges from analytical synthesis to complete architectural optimization. For this study, the consistent ABCD-based approach ensures that all topologies share the same NTF, quantizer resolution, and target SNR, enabling a fair comparison.

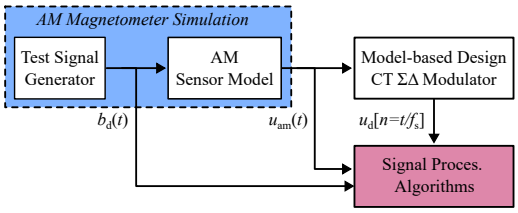


Fig. 2: Model-based evaluation flow: sinusoidal and synthetic MCG signals are amplitude-modulated, digitized by a CT ΣΔ modulator, and demodulated to recover the baseband signals.

### C. Model-Based Design and Evaluation Framework

Fig. 2 illustrates the unified evaluation flow. The AM magnetometer signal is synthesized, coefficients are derived using the ABCD framework, and CT ΣΔ behavioral models (CIFB, CIFFB, CIFF) are implemented in MATLAB/Simulink. The output is then demodulated and low-pass filtered to recover the baseband signal that was previously modulated onto the AM signal's envelope. All modulators operate with $f_\mathrm{s} = 12.8\,\mathrm{MHz}$ and $\mathrm{OSR} = 64$, corresponding to a ΣΔ bandwidth of $\approx 100\,\mathrm{kHz}$. A 15-level quantizer is used, where it provides a nominal in-band SNR of $97\,\mathrm{dB}$. These settings remain fixed to isolate the effect of topology. To characterize envelope-transfer behavior, a $20\,\mathrm{kHz}$ carrier is set for the AM signal with either a simulated sinusoidal wave (for $m$ parameter variation) or a synthetic MCG waveform (for simulation of a (bio-)magnetic signal. This enables direct comparison of linearity, noise-shaping behavior, and robustness across all three topologies, assuming an ideal magnetometer behaviour with no additional noise contributions. Signal comparison of the desired input signal $b_\mathrm{d}$, the amplitude-modulated $u_\mathrm{am}$, and the ΣΔ modulator output $u_\mathrm{d}$ is performed via baseband conversion. The corresponding signal-processing chain, implemented in MATLAB, consists of a Hilbert transform, offset compensation, and a third-order Butterworth low-pass filter (cutoff at $500\,\mathrm{Hz}$).

 DOI 10.15203/99106-194-6-18

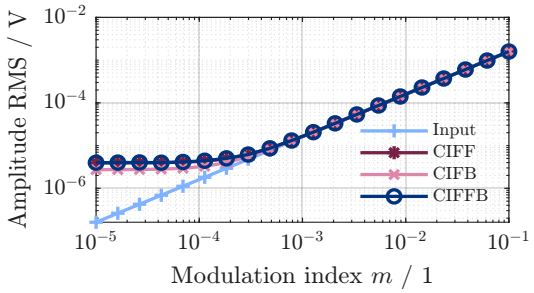


Fig. 3: Recovered sinusoidal signals at the baseband to calculate the output RMS versus modulation index for CIFB, CIFFB, and CIFF modulators, considering the noise-shaping overall noise.

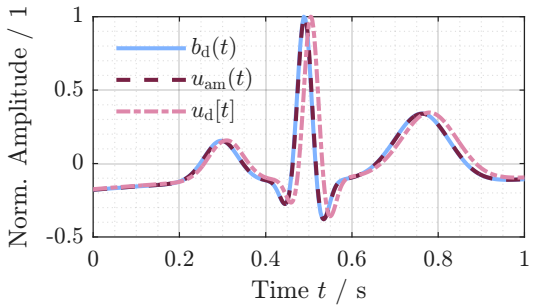


Fig. 4: Normalized baseband signals after reconstruction of a synthetic MCG waveform using a CIFFB $\Sigma\Delta$ modulator in comparison with the AM of $m = 0.001$ and the $\Sigma\Delta$ conversion in baseband.

## III. Results

Fig. 3, where the envelope test signal is $b_d(t) = \sin(2\,\pi\,f_b\,t)$ with $f_b = 19.44\,\text{Hz}$; the RMS is computed over 64 periods after discarding the first 10 periods to remove startup transients. For very small indices ($m \leq 10^{-3}$), all modulators approach the noise-shaping-limited RMS level. In the linear region ($10^{-3}$ to $10^{-1}$), the envelope scales proportionally with $m$, producing the expected $20\,\text{dB/dec}$ trend (amplitude scaling). The CIFB achieves the lowest RMS noise floor at small $m$, whereas CIFF and CIFFB exhibit a slightly elevated floor (approx $1.2\,\mu\text{V}$ higher). The $m \geq 0.1$ exceed the intended operating range of cME magnetometers and were therefore not considered further. Figure 4 shows the reconstructed envelope for a simulated MCG input with a $m = 0.001$. CIFFB is shown here because it offered the best envelope waveform fidelity (lowest visible distortion) in the synthetic MCG case to preserve the P–QRS–T morphology. A consistent delay is observed, determined by the combined $\Sigma\Delta$ dynamics and the subsequent baseband demodulation.

## IV. Discussion

The results show that the $\Sigma\Delta$ loop topology designed using the ABCD-based analytical approach helps for early-stage exploration and cross-topology comparison. It is important to note that these simulations were performed under noise-free conditions. Consequently, the reconstruction limit at low $m$ is determined solely by the quantization noise shaped by the $\Sigma\Delta$ loop. The comparison between the simulated MCG signal, the AM envelope, and the reconstructed envelope after $\Sigma\Delta$ conversion confirms that accurate baseband recovery is achievable even at ultra-low $m$. CIFFB offers the best trade-off between noise-shaping and envelope fidelity for the synthetic MCG waveform, even though CIFB achieves the lowest in-band noise at ultra-low $m$. The close alignment of the reconstructed envelope with the reference signals indicates that the loop preserves both amplitude and temporal features of the MCG waveform. Furthermore, the observed delay between the baseband AM envelope and the $\Sigma\Delta$-derived is consistent with the $\Sigma\Delta$ dynamics and the group delay of the demodulation/low-pass filtering chain and can be compensated in post-processing.

## V. Conclusion

This work (behavioral simulation) demonstrates that CT third-order $\Sigma\Delta$ topologies constitute suitable candidates for envelope extraction after baseband conversion in cME magnetometers operating at modulation indices below 0.1% ($m < 0.001$). Future work must account for circuit-level non-idealities such as thermal noise, jitter, finite gain, and excess-loop delay, and explore large-scale architecture optimization and additional noise sources, for example, those arising from the magnetometer. Incorporating these considerations is essential when transitioning from behavioral models to transistor-level implementations intended for integrated readout of an AM magnetometer used for (bio-)magnetic sensing.

## Acknowledgment

This work was funded by the DFG through project B1 of CRC 1261 *"Biomagnetic sensing"* (Grant 286471992).

## References

[1] E. Elzenheimer *et al.*, "Characterizing timing parameters in commercial SERF-OPM multichannel systems for biomagnetic field sensing," *Measurement*, vol. 263, 2026.
[2] H. Wang *et al.*, "Electronic Interface Design Considerations for Biomagnetic Sensing Using TMR Sensors," *IEEE TMAG*, 2025.
[3] T. P. Gutteling *et al.*, "A New Generation of OPM for High Dynamic and Large Bandwidth MEG: The 4He OPMs—First Applications in Healthy Volunteers," *Sensors*, vol. 23, no. 5, 2023.
[4] E. Elzenheimer, P. Hayes *et al.*, "Investigation of Converse Magnetoelectric Thin-Film Sensors for Magnetocardiography," *IEEE Sensors J.*, 2023.
[5] S. Pavan *et al.*, *Understanding Delta-Sigma Data Converters*, 2nd ed. Wiley-IEEE Press, 2017.
[6] M. Cortez *et al.*, "Behavioral and Electrical Modeling of a 0.5-V Third-Order Continuous-Time Sigma-Delta Modulator with FIR DAC for Audio Applications," *J. of Integrated Circ. and Syst.*, no. 1, May 2023.
[7] J. M. de la Rosa and R. del Río, *CMOS Sigma-Delta Converters: A Practical Design Guide*. Wiley, 2013.
[8] J. Solis-Arbustini *et al.*, "A 16-bit Pressure Sensing Interface Integrating a 460 fJ/conv Incremental Sigma Delta ADC for Medical Devices," in *IEEE LASCAS Proc.*, 2020.
[9] J. Arbustini *et al.*, "Analog Demodulation for Electrically Modulated ME Sensors: A Signal Integrity Analysis and Evaluation Approach," *IEEE Sensors J.*, 2025.
[10] J. Wagner, P. Vogelmann, and M. Ortmanns, "An Automated Design Environment for CT Incremental Sigma-Delta ADCs," in *IEEE ICECS Proc.*, 2020.

 DOI 10.15203/99106-194-6-19

# Radial Harmonic Fourier Moments for Radiogenomics

A.H.M. Sajedul Hoque, Gergő Bognár, Sándor Fridli

***Abstract*—Cancer treatment relies on understanding the biological pathways that drive tumor progression, yet traditional biopsies are invasive and capture only a small portion of a heterogeneous tumor. Radiomics provides a non-invasive alternative by extracting quantitative features from CT images that may reflect underlying molecular processes. Texture features are particularly important for characterizing tumor heterogeneity, but widely used wavelet-based features are computed in the Cartesian domain and are sensitive to rotation, translation, and scaling, limiting their robustness.**

**To address these limitations, we investigated Radial Harmonic Fourier Moments (RHFMs), a polar-frequency representation that is inherently invariant to geometric transformations and captures both global and local texture information with fewer coefficients. By integrating RHFM-derived radiomic features with genomic data, we examined their associations with biological pathways in lung cancer. Our results show that RHFMs yield more stable and biologically meaningful radiogenomic correlations than wavelet features, highlighting their potential as reliable non-invasive biomarkers.**

***Index Terms*—radiomics, radiogenomics, lung CT, orthogonal moments, RHFM, GSEA**

## I. Introduction

Cancer is fundamentally a genetic disease driven by mutations that disrupt critical cellular pathways, leading to hallmark behaviors such as uncontrolled proliferation, invasion, and metastasis [1]. Identifying the biological processes underlying these behaviors is essential for developing effective and targeted therapies.

Tumor biology is traditionally assessed using tissue biopsies; however, biopsies are invasive, costly, and sample only limited tumor regions, failing to capture spatial and temporal heterogeneity. In contrast, medical imaging—particularly X-ray computed tomography (CT)—provides a non-invasive alternative. Radiomics enables the extraction of quantitative imaging features that may reflect underlying tumor biology [2]. In [3], radiomic features are categorized into shape, intensity, second-order texture, and higher-order texture descriptors. Among these, wavelet-based features are widely used to capture multi-scale frequency information, but their Cartesian formulation makes them sensitive to geometric transformations, limiting robustness and reproducibility.

A.H.M. Sajedul Hoque is with the Department of Numerical Analysis, ELTE Eötvös Loránd University, Hungary, e-mail: hoque.cse@cu.ac.bd

Gergő Bognár and Sándor Fridli are with the Department of Numerical Analysis, ELTE Eötvös Loránd University, Hungary.

Project no. K146721 has been implemented with the support provided by the Ministry of Culture and Innovation of Hungary from the National Research, Development and Innovation Fund, financed under the K_23 "OTKA" funding schemes. This work was supported by the University Excellence Fund of Eötvös Loránd University, Budapest, Hungary (ELTE).

To overcome these limitations, we proposed radiomic features by means of Radial Harmonic Fourier Moments (RHFMs) in our previous work [4]. Image moments are widely used numerical descriptors of shape, texture, and frequency content, offering sparse and geometric invariant representation (see e.g. [5] for a recent survey). In particular, RHFMs [6] provide a compact polar-frequency representation that is inherently invariant to translation, rotation, and scaling. Motivated by these advantages compared to wavelets, we extracted higher-order radiomic features using RHFMs in [4], and demonstrated superior repeatability and comparable prognostic performance relative to wavelet features, suggesting their potential as a more stable radiomic alternative. However, it remains unclear whether these advantages translate into stronger and more biologically meaningful associations with cancer-related molecular pathways.

This study addresses this gap by evaluating the biological relevance of RHFM-based radiomic features using radiogenomics analysis. To improve interpretability and reduce redundancy, we distilled a high-dimensional radiomic feature space (436 features) into a compact four-feature radiomic signature, selecting one robust and prognostic feature from each major radiomic category. The central objective of this work is to determine whether this stable radiomic signature—particularly incorporating RHFM-derived features—can reliably uncover biologically meaningful pathways associated with lung cancer.

## II. Materials and Methods

The proposed workflow (Fig. 1) comprises three main phases: (1) Radiomic Feature Analysis, (2) Rad-Score Computation, and (3) Radiogenomic Association Analysis.

Radiomic features are first extracted and assessed for repeatability, robustness, and prognostic relevance following our previous methodology [4]. A compact radiomic signature is then constructed and used to compute patient-specific Rad-Scores. Finally, these imaging-derived phenotypes are linked to gene expression profiles to identify biologically relevant pathways associated with lung cancer.

### A. Datasets and Radiomic Features

Four publicly available lung cancer cohorts described in [3] were utilized: RIDER Test–Retest (27 patients) for repeatability assessment, Multiple Delineation (21 patients) for inter-observer robustness, Lung1 (421 patients) for prognostic evaluation using the concordance index (CI), and Lung3 (89

 DOI 10.15203/99106-194-6-19

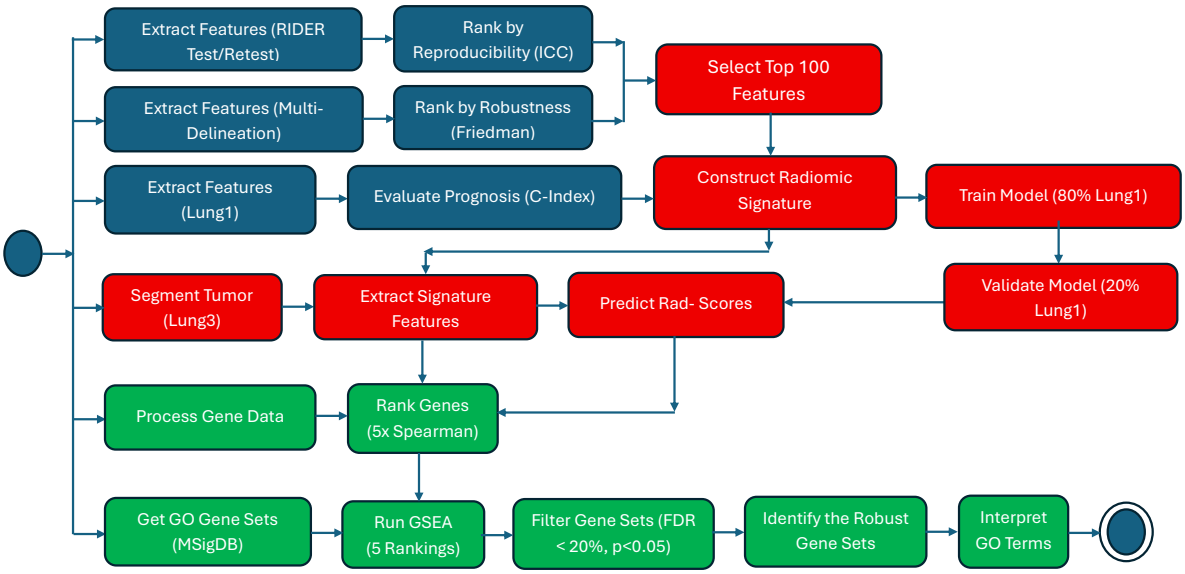


Fig. 1: Overview of the proposed radiogenomic workflow. Main phases: Radiomic Feature Analysis (blue), Rad-Score Computation (red), Radiogenomic Association Analysis (green)

patients) for radiogenomic association analysis using matched CT and gene expression data.

Radiomics transforms CT images into quantitative descriptors of tumor shape, intensity, and texture, reflecting underlying biological heterogeneity. Conventional radiomic descriptors provide multi-scale representations that capture complementary spatial and frequency-domain tumor characteristics. Using PyRadiomics [7] and following the framework of [3] features were extracted across five categories: 15 first-order statistics (FOS), 13 shape features, 21 Gray-Level Co-occurrence Matrix (GLCM) features, 11 Gray-Level Run-Length Matrix (GLRLM) features, and 376 higher-order features from either RHFM or wavelet representations. The main focus of this work is on the RHFM features following our prior work [4], where we reconstructed CT images using eight different RHFM orders. For comparison purposes, we also investigated wavelet features proposed as of [3], where CT volumes are decomposed into eight wavelet subbands (using coiflets). We note that the number of the utilized radiomic features varies in the literature. Here, we worked with 436 robust and reproducible features only, after excluding redundant or ill-defined descriptors used in [3] and [4].

### B. Rad-Score Computation

Radiomic features were ranked using the Intra-Class Correlation Coefficient (ICC) for test–retest repeatability and the Friedman test $p$-value for inter-observer robustness [4]. The average rank of both metrics was used to select the top 100 most stable features. Each feature was then assessed for prognostic significance in the Lung1 cohort using a univariate Cox proportional hazards model as of [4]. Then, the highest-concordance feature from each radiomic category was selected to form a compact four-feature radiomic signature following the guidelines of [3].

The Lung3 dataset includes both CT images and genome-wide gene expression profiles (22,115 genes across 89 patients). Following tumor segmentation, the four signature features were extracted, yielding feature vectors aligned with patient-level gene expression data.

A multivariate CoxPH model was trained on Lung1 using an 80/20 split to estimate feature contributions. The resulting coefficients ($\beta_j$) were applied to Lung3 features to compute patient-specific Rad-Scores:

$$\text{Rad-Score}_i = \sum_{j=1}^{p} \beta_j X_{ij}, \tag{1}$$

where $X_{ij}$ denotes the value of feature $j$ for patient $i$. The Rad-Score provides a single prognostic imaging phenotype suitable for radiogenomic analysis.

### C. Radiogenomic Association Analysis

Genes were ranked using Spearman correlation in five ways: one ranked list based on the overall Rad-Score and four additional lists based on Rad-Scores computed from each individual feature. Pre-ranked Gene Set Enrichment Analysis (GSEA) [8] was applied to each list using the Gene Ontology Biological Process (C5) collection from MSigDB [9]. GSEA outputs included the Enrichment Score (ES), Normalized Enrichment Score (NES), $p$-value, and False Discovery Rate (FDR). Gene sets with $FDR < 0.20$ and $p < 0.05$ were considered significant. Gene sets were defined as robust if they appeared in both the overall Rad-Score analysis and at least one feature-specific analysis. These robust Gene Ontology (GO) terms were subsequently interpreted from an oncological perspective to identify cancer-relevant biological processes.

## III. Results and Discussion

This section presents the construction of compact radiomic signatures, their prognostic performance, and the biological interpretation of associated gene expression patterns.

### A. Radiomic Signature Construction

To improve interpretability and reduce redundancy, one robust and prognostic feature from each radiomic category was selected to form compact four-feature radiomic signatures. Tables I and II summarize the wavelet- and RHFM-based signatures, respectively.

TABLE I: Wavelet-Based Radiomic Signature.

| Group | Feature | CI |
|---|---|---|
| FOS | Uniformity | 0.58 |
| Shape | Flatness | 0.50 |
| Texture (GLRLM) | Gray Level Non-Uniformity | 0.58 |
| Wavelet (HLH) | IMC2 | 0.62 |

TABLE II: RHFM-Based Radiomic Signature.

| Group | Feature | CI |
|---|---|---|
| FOS | Uniformity | 0.58 |
| Shape | Surface Volume Ratio | 0.60 |
| Texture (GLRLM) | Gray Level Non-Uniformity | 0.58 |
| RHFM (Order 2) | Median | 0.62 |

Both signatures capture complementary tumor characteristics; however, the RHFM-based signature shows improved shape-related prognostic relevance.

### B. Prognostic Performance

Multivariate Cox proportional hazards models were trained on the Lung1 dataset using each signature. Performance comparison is reported in Table III.

TABLE III: Performance Comparison of Prognostic Models.

| Metric | RHFM Model | Wavelet Model |
|---|---|---|
| Test C-index | 0.60 | 0.58 |
| Bootstrap C-index | $0.59 \pm 0.03$ | $0.58 \pm 0.02$ |

The RHFM-derived model achieved similar, but slightly higher concordance indices, indicating improved stability and discrimination compared to the wavelet-based model.

### C. Radiogenomic Associations

Pre-ranked GSEA was performed using the MSigDB C5 collection. After filtering gene sets as of [3] (15–500 genes), 8,604 gene sets were analyzed, that yielded 1,127 significantly enriched GO terms. We considered the top 20 most significant upregulated and top 20 downregulated gene sets for detailed analysis. Enriched gene sets were mapped to the ten cancer hallmarks defined in [1]. Table IV summarizes hallmark distributions for the 40 most significant gene sets, covering the affected hallmarks only, such as sustaining proliferative signaling (Prolif.), deregulating cellular energetics (Metab.), avoiding immune destruction, genome instability & mutation, tumor-promoting inflammation (Inflam.).

TABLE IV: Hallmark Distribution Associated with RHFM and Wavelet Features.

| Type | Cancer Hallmarks | | | | | Other |
|---|---|---|---|---|---|---|
| | Prolif. | Metab. | Immune | Genomic | Inflam. | |
| Wavelet | – | 5% | 2.5% | 7.5% | 5% | 80% |
| RHFM | 35% | – | 35% | – | 2.5% | 27.5% |

RHFM-derived features exhibit strong and biologically coherent enrichment, reflecting key tumor cell-cycle and immune-regulatory mechanisms. In contrast, wavelet-derived features show weak and largely non-specific biological associations among their most significant up- and downregulated processes, with most gene sets falling outside canonical cancer hallmarks.

## IV. Conclusion

In this study, we proposed a radiogenomic framework using RHFM-based radiomics, and compared the results to state-of-the-art wavelet-based approach. We evaluated their prognostic performance and their ability to uncover biologically meaningful associations with gene expression in lung cancer.

Our results demonstrate that RHFM-derived radiomic signatures yield more stable prognostic models and show stronger and more coherent enrichment of cancer-related biological pathways, particularly those associated with proliferative signaling and immune-related processes. This indicates that RHFMs capture tumor heterogeneity in a manner that is not only robust and prognostically relevant but also closely aligned with underlying molecular mechanisms.

Overall, this study highlights the potential of RHFM-based radiomics as a biologically grounded and clinically meaningful bridge between medical imaging and tumor genomics, supporting more interpretable, reliable, and personalized cancer assessment in precision oncology.

## References

[1] D. Hanahan and R. A. Weinberg, "Hallmarks of cancer: The next generation," *Cell*, vol. 144, no. 5, pp. 646–674, Mar. 2011.

[2] G. Lin *et al.*, "Radiomic Models Predict Tumor Microenvironment Using Artificial Intelligence-the Novel Biomarkers in Breast Cancer Immune Microenvironment," *Technology in Cancer Research & Treatment*, vol. 22, p. 15330338231218227, 2023.

[3] H. Aerts *et al.*, "Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach," *Nature Communications*, vol. 5, p. 4006, Aug 2014.

[4] A. H. M. S. Hoque *et al.*, "Radial harmonic Fourier moments for CT-based quantitative radiomics," *Acta Cybernetica*, vol. 27, no. 2, p. 175–196, Mar. 2025.

[5] S. Qi *et al.*, "A survey of orthogonal moments for image representation: Theory, implementation, and evaluation," *ACM Comput. Surv.*, vol. 55, no. 1, Nov. 2021.

[6] Y. Liu *et al.*, "Accurate quaternion radial harmonic fourier moments for color image reconstruction and object recognition," *Pattern Analysis and Applications*, vol. 23, pp. 1–17, 11 2020.

[7] J. J. M. Van Griethuysen *et al.*, "Computational Radiomics System to Decode the Radiographic Phenotype," *Cancer Research*, vol. 77, no. 21, pp. e104–e107, Nov. 2017.

[8] A. Subramanian *et al.*, "Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 102, no. 43, pp. 15 545–15 550, 2005.

[9] A. Liberzon *et al.*, "Molecular signatures database (MSigDB) 3.0," *Bioinformatics*, vol. 27, no. 12, pp. 1739–1740, 05 2011.

 DOI 10.15203/99106-194-6-20

# Joint Temporal and Feature Level Attention Patterns for Interpretable Sepsis Prediction from Unreliable Multivariate ICU Patient Data

Daniel Breucker, M.Sc., Kai Wu, M.Sc., Johannes Krefting, MD, Karsten Wenger, Dr., Shweta Mahajan, M.Sc, Irina Pugach, Dr., Daniela Ramirez, B.Sc, Marius Schwab, MD, Christian Frieß, MD, Moritz von Scheidt, MD

***Abstract*—Sepsis remains a major concern in Intensive Care Units (ICUs), where timely intervention is critical, yet clinicians still rely on rule-based bedside Early Warning Systems (EWS) that lack automation and real-time monitoring. Despite the growing availability of clinical data, its unreliable, multivariate, and heterogeneously sampled nature complicates deep learning applications. Although recent models surpass clinical gold standards, their limited interpretability restricts clinical adoption. This work aims to improve interpretability in Sepsis-3 onset prediction using such complex ICU data. We propose an attention-based framework that provides joint temporal and feature-level explanations, supported by a tailored dataset derived from the eICU collaborative research database and an adapted Gated Transformer Network (GTN) to incorporate static patient information. Combined attention patterns are used to visualize the model's decision-making process. Our results show that Sepsis-3 onset can be predicted while simultaneously offering clinically meaningful interpretability, supporting the bench-to-bedside translation of deep learning in critical care.**

***Index Terms*—Deep Learning, Time Series, Interpretability, ICU Patient Data, Physiological Monitoring**

## I. Introduction

*1) Severity of Sepsis:* Sepsis is a substantial source or rather a sink of Years Life Lost (YLL), occupying the 12th position in the Global Burden Survey of 2010, close to HIV and Malaria (5th and 6th rank respectively) [1].

It is a severe condition for which no targeted therapy is available but treatable, albeit under substantial time pressure [2], [3].

*2) Overwhelming Patient Data:* Clinical healthcare is increasingly adopting connected medical systems similar to industrial IIoT, creating opportunities to streamline workflows and improve patient outcomes [4]. Realizing these benefits, however, depends on the ability to process and analyze the large volumes of data such systems generate. Current clinical practice is not fit for this data influx: existing EWSs rely on limited subsets of available information and provide only snapshot assessments, failing to capture the full trajectory of patient deterioration [5].

D. Breucker, K. Wu, J. Krefting, K. Wenger, S. Mahajan, I. Pugach, D. Ramirez, M. Schwab, C. Frieß, M. von Scheidt are with the Department of Cardiology, TUM University Hospital German Heart Center, TUM School of Medicine and Health, Technical University of Munich, Munich, Germany.

D. Breucker is with the Heinz Nixdorf Chair of Biomedical Electronics, TUM School of Computation, Information and Technology, Technical University of Munich, Munich, Germany.

*3) Contributions:* Our contributions amount to Sepsis-3 onset prediction with unreliable multi-channel heterogeneously sampled ICU patient data with a dense recording interval of $5\,\mathrm{min}$ while also providing a sample-wise joint temporal and feature level ante-hoc interpretability method via the attention mechanism for clinical decision making.

## II. Methods

*1) Approach:* Our approach consists of following core components, we develop a deep learning model featuring the attention mechanism for Sepsis-3 onset prediction in ICU patients using the eICU dataset. The key challenge in this context is the unreliable multi-variate heterogeneously sampled patient data, where some features, like physiological vital signs, are densely populated, while others, such as lab tests, are sparsely available or even absent despite their critical clinical importance and others are completely static like patient demographics or comorbidities. Additionally, we leverage the attention mechanism inherent to our chosen model to explore its decision-making process. By visualizing temporal and feature-level attention patterns, we can determine whether the model focuses on clinically relevant temporal patterns or features.

*2) Model:* Our primary model is the GTN [6], that we adapt for the inclusion of static inputs. This model is well-suited for our interpretability objective due to its reliance on attention layers, whose scores can be examined to illustrate the model's decision-making process. The GTN consists of two transformer encoder blocks: one operating along the temporal dimension to model sequence dependencies, and the other along the feature dimension to capture interactions between individual variables, independent of time. This design enables joint temporal and feature-wise analysis, supporting attention-based visualization.

*3) Interpretability:* We facilitate an ante-hoc method taking advantage of the attention mechanism [7].

$$\mathrm{Attention}(\mathbf{Q},\mathbf{K},\mathbf{V}) = \underbrace{\mathrm{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^\top}{\sqrt{d_k}}\right)}_{\text{attention pattern (AP)}}\mathbf{V}\,, \qquad (1)$$

To leverage the attention mechanism for interpretability, we calculate the attention score, but we will refer to it in subsequent sections as the attention pattern for each prediction of interest, as the visualization of that yields distinct patterns,

as described in Equation 1. By combining the temporal and feature-level attention patterns via the outer product, as shown in Equation 3, we create a compact visualization of the model's decision-making process. We calculate the combined attention pattern $\mathrm{AP}_{\mathrm{combined}}$ as,

$$\mathbf{AP}_{\mathrm{combined}} = \overline{\mathrm{AP}}_{\mathrm{temporal}} \otimes \overline{\mathrm{AP}}_{\mathrm{features}} \tag{2}$$

$$= \begin{array}{c} \\ t_1 \\ t_2 \\ \vdots \\ t_T \end{array} \begin{array}{c} \begin{array}{cccc} f_1 & f_2 & \cdots & f_F \end{array} \\ \begin{bmatrix} \overline{\alpha}_1\overline{\beta_1} & \overline{\alpha}_1\overline{\beta_2} & \cdots & \overline{\alpha}_1\overline{\beta_F} \\ \overline{\alpha}_2\overline{\beta_1} & \overline{\alpha}_2\overline{\beta_2} & \cdots & \overline{\alpha}_2\overline{\beta_F} \\ \vdots & \vdots & \ddots & \vdots \\ \overline{\alpha}_T\overline{\beta_1} & \overline{\alpha}_T\overline{\beta_2} & \cdots & \overline{\alpha}_T\overline{\beta_F} \end{bmatrix} \end{array}, \tag{3}$$

with $\mathrm{AP}_{\mathrm{combined}} \in \mathbf{R}^{T \times F}$ being the averaged feature attention pattern weighted by the averaged temporal attention at each time step. This visualization allows us to investigate to which points in time the model attends in conjunction with features.

*4) Data Preprocessing:* Implausible or incorrectly scaled values were removed or converted, and physiologically infeasible measurements (e.g., body temperatures $> 48$ °C or negative blood pressures) were discarded, particularly among periodic and aperiodic variables. After manual thresholding, we applied percentile-based outlier removal: ±3 $\sigma$ for periodic vital signs and the 0.001–99.99 percentile range for aperiodic variables. Laboratory values were exempt from percentile filtering due to their sparsity and generally higher reliability. Cleaned tables were then merged, and features with fewer than 15,000 data points were removed to eliminate rarely populated variables. Time series were aligned to a common 5-minute grid and segmented into non-overlapping 6-hour chunks, with insufficiently long segments discarded. To avoid samples dominated by missingness introduced during alignment, we excluded patient stays with a heart rate density below 0.4, using heart rate as a proxy for overall data availability given its prevalence in ICU monitoring. The dataset was stratified by sepsis status, determined hourly using the RICU package [8], to ensure balanced distributions of case and control patients. Scaling parameters for Min–Max normalization were derived from the stratified training set and applied universally, normalizing-to-Gaussian methods were avoided as they are not justified for most variables. Static categorical features were encoded accordingly. Because each patient may generate multiple negative chunks but at most one positive sample, the resulting dataset was highly imbalanced. We therefore performed Random Under Sampling to approximate a sepsis prevalence of 30%, consistent with ICU Sepsis-3 estimates [9], reducing excess negative samples.

*5) Experimental Setup:* Our task is to predict Sepsis-3 onset using non-overlapping 6-hour input windows, paired with a 6-hour prediction horizon. The sliding window begins at ICU admission and moves forward through each patient's time series. Sepsis onset is retrospectively determined using the RICU package [8]. A sample is labeled positive if Sepsis-3 onset occurs within the subsequent prediction horizon; sepsis patients therefore contribute multiple negative samples prior to meeting the criteria and a single positive sample at onset. To align labels with the input window, we shift them back by the length of the prediction horizon, enabling binary classification on the available samples. The dataset is split 60:20:20 into training, validation, and testing sets. Each sample comprises a time-series component of size $T \times TF$ (72 × 69) and a static feature vector of size $1 \times SF$ (1 × 12). Class imbalance is handled during training via weighted random sampling with undersampling of the majority class, sampling is disabled during validation and testing to reflect the assumed sepsis prevalence. Before random undersampling, the dataset comprised 122,831 training samples (456 sepsis, 122,375 control), 40,944 validation samples (152 sepsis, 40,792 control), and an identical 40,944 samples in the test set. After undersampling to adjust class imbalance, the training set was reduced to 1,976 samples (456 sepsis, 1,520 control), with both the validation and test sets containing 658 samples each (152 sepsis and 506 control). We evaluate three models: the GTN as the primary architecture and Simply Attend and Diagnose (SAnD) and Long Short-Term Memory (LSTM) as baselines. Training uses binary cross-entropy with logits loss and the AdamW optimizer, with learning rate and weight decay selected through hyperparameter search. Learning rate reduction on plateau and early stopping are applied based on validation performance. Cross-validation was not performed due to limited dataset size and the focus on methodological rather than clinical validation. We perform hyperparameter tuning to find an optimal configuration for the given experimental setup, we employed a Bayesian search with Hyperband early termination ($s = 2$, $\eta = 3$, $max_{iter} = 20$). AUROC on the validation set across 200 individual trials was the optimization task. Resulting in the following model configurations. **GTN**: Time-series ($72 \rightarrow 32$) and feature inputs ($69 \rightarrow 32$) are embedded into 32-D space, processed by single-layer 8-head self-attention with a FFN ($32 \rightarrow 128 \rightarrow 32$), gated across channels, compressed via a bottleneck ($32 \rightarrow 3$), and concatenated with static features ($12 \rightarrow 128 \rightarrow 12$) for prediction via a FC classifier. **SAnD**: Time series are embedded to 32-D using 1-D convolution and single-layer 8-head self-attention, temporally aggregated via dense interpolation (factor 36), concatenated with static embeddings ($12 \rightarrow 32$), and classified using a single FC head. **LSTM**: Time series are encoded with a 2-layer LSTM (hidden size 128) followed by projection ($128 \rightarrow 8$), static features are embedded separately, concatenated, and passed through a FC classifier.

## III. Results

While performance was not the primary focus of this thesis, the performance evaluation of the models provides an essential backdrop for the interpretability assessment of the decision-making processes. As shown in Figure 1, all models exhibit limited performance. However, the LSTM performs so poorly that it suggests the model may not be capable of providing meaningful predictions. This outcome is mainly attributable to the experimental setup and model architectures of GTN, SAnD, and LSTM.

Figure 2 shows the combined attention pattern for a specific sample. We generate this attention pattern by calculating the

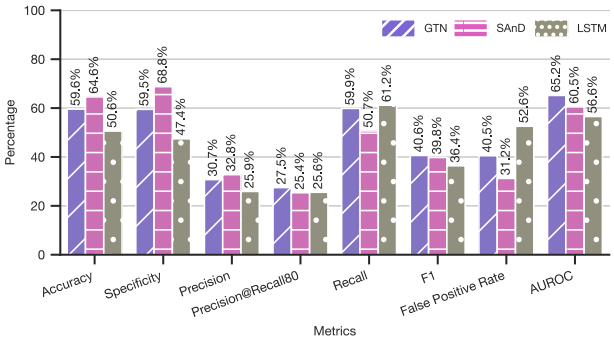


Fig. 1: Scalar performance metrics of the GTN, SAnD, LSTM models on the test set in our experiment. Note: The proposed approach prioritizes clinical interpretability over raw performance.

outer product of the query-wise attention pattern averages see Equation 3, effectively joining the temporal and feature-level attention patterns for all available features in a compact manner. This approach might be more suitable for clinical environments, where the isolated temporal or feature-level attention patterns may not be as informative as their combined representation. In this visualization, dense vital signs like SaO2 and respiration remain prominent, but lab tests such as Lymphocytes, which are essential for assessing the patient's immunological state, and Total Bilirubin, which is used in the Sequential (sepsis-related) Organ Failure Assessment (SOFA) score, are also more easily distinguished as important features according to the attention scores.

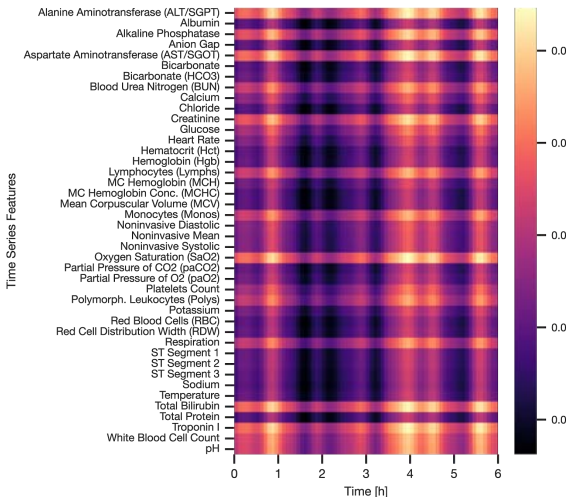


Fig. 2: Combined attention pattern which joins temporal and feature-level attention for a sample, which we calculate via Equation 3. The temporal attention score changes at each time step, while the feature-wise attention score changes across features but not over time.

## IV. Discussion

Model performance directly shapes the quality of interpretability results. Although we visualize the model's decision-making, these explanations may not fully reflect competent reasoning due to performance limitations. Still, providing interpretability, even if imperfect, remains preferable to offering none, consistent with clinicians preference for transparency over raw accuracy [10]. Our attention-based interpretability method is conceptually independent of predictive performance, it can illustrate why a prediction is likely correct or incorrect. Further work is needed to understand the limits of attention-driven explanations, especially with respect to sparse versus dense time series features. An ablation study could clarify whether meaningful benefits arise from these mechanisms. Leveraging the data warehouse at our center (TUM University Hospital German Heart Center) may provide means for extended investigation of the approach. Future work should also address trustworthiness, stability, and robustness, properties essential for ICU deployment.

## V. Conclusion

We presented a method for predicting Sepsis-3 onset 6 hours in advance using a Gated Transformer Network model applied to unreliable, multivariate, and heterogeneously sampled ICU data, including vital signs, laboratory tests, comorbidities, and demographics. By leveraging attention mechanisms, the approach visualizes intra-sample temporal and feature interactions, exemplified through combined attention patterns.

Our results show that deep learning can be applied to challenging ICU data with minimal feature engineering, without handcrafted features or extensive aggregation, even at a 5-minute sampling interval. Moreover, the method provides interpretability suitable for clinical settings, addressing a key barrier to deploying deep learning–based early warning systems. These contributions offer a foundation for further research in clinically interpretable sepsis prediction.

## References

[1] R. Lozano, M. Naghavi, K. Foreman, *et al.*, "Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: A systematic analysis for the Global Burden of Disease Study 2010," *The Lancet*, vol. 380, no. 9859, pp. 2095–2128, Dec. 15, 2012.

[2] J. D. Faix, "Biomarkers of sepsis," *Critical Reviews in Clinical Laboratory Sciences*, vol. 50, no. 1, pp. 23–36, 2013.

[3] D. Jarczak, S. Kluge, and A. Nierhaus, "Sepsis—Pathophysiology and Therapeutic Concepts," *Frontiers in Medicine*, vol. 8, May 14, 2021.

[4] Z.-x. Lu, P. Qian, D. Bi, *et al.*, "Application of AI and IoT in Clinical Medicine: Summary and Challenges," *Current Medical Science*, vol. 41, no. 6, pp. 1134–1150, Dec. 1, 2021.

[5] R. C. Bone, R. A. Balk, F. B. Cerra, *et al.*, "Definitions for Sepsis and Organ Failure and Guidelines for the Use of Innovative Therapies in Sepsis," *Chest*, vol. 101, no. 6, pp. 1644–1655, Jun. 1, 1992.

[6] M. Liu, S. Ren, S. Ma, *et al.* "Gated Transformer Networks for Multivariate Time Series Classification." (Mar. 26, 2021), [Online]. Available: http://arxiv.org/abs/2103.14438, pre-published.

[7] T. Rojat, R. Puget, D. Filliat, J. Del Ser, R. Gelin, and N. Díaz-Rodríguez. "Explainable Artificial Intelligence (XAI) on TimeSeries Data: A Survey." (Apr. 2, 2021), [Online]. Available: http://arxiv.org/abs/2104.00950, pre-published.

[8] N. Bennett, D. Plečko, I.-F. Ukor, N. Meinshausen, and P. Bühlmann, "Ricu: R's interface to intensive care data," *GigaScience*, vol. 12, giad041, Dec. 28, 2022.

[9] "Sepsis is underreported in Swedish intensive care units: A retrospective observational multicentre study - Lengquist - 2020 - Acta Anaesthesiologica Scandinavica - Wiley Online Library." (), [Online]. Available: https://onlinelibrary.wiley.com/doi/full/10.1111/aas.13647.

[10] S. Tonekaboni, S. Joshi, M. D. McCradden, and A. Goldenberg, "What Clinicians Want: Contextualizing Explainable Machine Learning for Clinical End Use," in *Proceedings of the 4th Machine Learning for Healthcare Conference*, PMLR, Oct. 28, 2019, pp. 359–380.

 DOI 10.15203/99106-194-6-21

# Are IPUs Viable for Genomic Sequence Alignment? Performance and Memory Analysis

S.-Kazem Shekofteh, Nils Kochendörfer, Holger Fröning

***Abstract*—Modern genomic sequencing generates massive datasets requiring efficient sequence alignment algorithms. We investigate the Graphcore Intelligence Processing Unit (IPU), an AI-optimized accelerator, for the Needleman-Wunsch sequence alignment algorithm. GCUPS (Giga Cell Updates Per Second) measures the throughput of computing alignment matrix cells. Through comprehensive benchmarking on GC200 IPU hardware with 1,472 tiles and 918 MB on-chip memory, we demonstrate peak performance of 36 GCUPS for optimal diagonal computations—2.4× faster than NVIDIA A30 GPU baseline (15 GCUPS). However, memory constraints limit practical applications to sequences under 9,400 nucleotides versus 55,000 for GPU. We identify critical memory-performance trade-offs and provide practical guidelines for integrating IPUs into genomic pipelines. Our findings demonstrate IPUs complement rather than replace GPUs in bioinformatics infrastructure.**

***Index Terms*—DNA Sequence Alignment, Genomic Data Processing, High-Performance Computing, Intelligence Processing Unit, Needleman-Wunsch Algorithm, Bioinformatics**

## I. Introduction

### A. Motivation and Background

The exponential growth of genomic sequencing data demands novel computational approaches beyond traditional CPU and GPU architectures. Human chromosomes contain 50-260 million nucleotides, with the complete genome spanning approximately 3 billion base pairs. Aligning these sequences against reference databases requires massive computational resources, with the Needleman-Wunsch (NW) algorithm [5] serving as a cornerstone for global sequence alignment.

While GPUs have dominated bioinformatics acceleration [2], emerging AI-optimized hardware presents unexplored opportunities. The Graphcore Intelligence Processing Unit (IPU) features 1,472 independent processing tiles with unique architectural characteristics: each tile has 624 KB local memory and executes independently without GPU-style lockstep constraints. Unlike GPUs operating in SIMT mode, IPU tiles enable flexible handling of irregular workloads. This work investigates whether such AI hardware can effectively accelerate sequence alignment—a fundamental operation in genomic analysis pipelines.

### B. Research Questions (RQ)

We address three key questions: (1) Can IPUs achieve competitive performance with GPUs for sequence alignment? (2) What are the fundamental architectural constraints limiting genomic-scale processing? (3) For which specific bioinformatics scenarios are IPUs viable accelerators?

S.-Kazem Shekofteh, N. Kochendörfer, and H. Fröning are with the Hardware and Artificial Intelligence (HAWAII) Lab, Heidelberg University, Germany, e-mail: kazem.shekofteh@ziti.uni-heidelberg.de.

Graphcore IPU architectural details: [3].

## II. Methods

### A. IPU Architecture and Poplar SDK

The Graphcore GC200 IPU provides 918 MB total on-chip memory distributed across 1,472 tiles, each supporting up to six parallel threads. The exchange fabric provides 350 GB/s inter-chip bandwidth and 46.9 TB/s memory bandwidth, with 62.5 TFlops/s FP32 peak performance at 150W power consumption. Programming uses the Poplar SDK with explicit graph construction and static memory allocation at compile time, following the Bulk-Synchronous Parallel model. The SDK provides configuration parameters for optimization strategy deployment: vertex granularity (elements per compute unit), frame sizing (consecutive elements per vertex), and workload distribution schemes—all analyzed in our experiments.

### B. Algorithm Implementation

The Needleman-Wunsch algorithm computes an alignment score matrix $M$ where:

$$M(i,j) = \max \begin{cases} M(i-1,j-1) + \omega(a_i, b_j) \\ M(i-1,j) + \delta \\ M(i,j-1) + \delta \end{cases} \quad (1)$$

with $\omega$ as the match/mismatch scoring function and $\delta$ as the gap penalty. We exploited diagonal wavefront parallelization where each diagonal $d$ contains cells computable in parallel.

We implemented a three-phase strategy: (1) **Build-up phase**: diagonals grow from 1 to 8,832 elements (1,472 tiles × 6 threads); (2) **Saturated phase**: maintains 8,832 threads with dynamic workload distribution; (3) **Culmination phase**: diagonals shrink back.

### C. Novel Optimization Techniques

Our key contribution is "frame sizing"—vertices (tile programs) computing multiple consecutive diagonal elements rather than single elements. This novel optimization dramatically reduces memory overhead by decreasing the number of vertex instances and control structures. While diagonal parallelization is standard [2], our frame sizing approach specifically addresses IPU's static compilation constraints and is absent from prior GPU implementations.

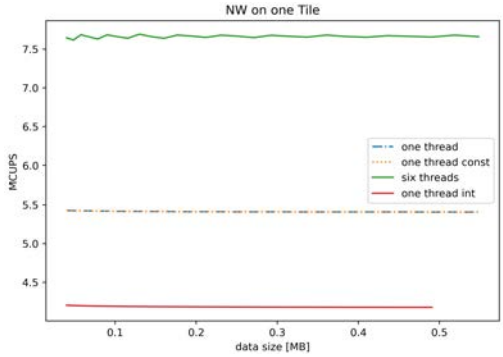


Fig. 1: Single-tile performance: integer (4.25 MCUPS), single-thread FP (5.5 MCUPS), six-thread FP (7.5 MCUPS).

Three critical optimizations were employed: **Vertex data reduction** (our contribution) minimizing per-vertex memory—reducing passed variables from 7 to 1 decreased exchange code by 45% and vertex instances by 60%. **Frame size optimization** (our contribution) created memory-performance trade-offs where frame=1 maximizes performance while frame=50 maximizes capacity. **Dynamic workload distribution** (adapted from [1]) maintained full tile utilization by adaptively increasing work per thread as diagonal length grows.

### D. Experimental Setup

Benchmarks used Graphcore M2000 system (one of four GC200 IPUs) with NVIDIA A30 GPU baseline (165W power, comparable to GC200's 150W). Synthetic DNA sequences (generated using Python's random module with uniform nucleotide distribution) ranged from 500 to 36 million elements. Performance measured in GCUPS (Giga Cell Updates Per Second). Memory profiling employed PopVision Graph Analyzer capturing control code, vertex code, vertex instances, exchange code, and variable memory usage. Integer computations used int32 precision; floating-point used FP32 throughout.

## III. Results

Our experiments systematically evaluate IPU performance across different parallelization strategies and compare against GPU baseline to answer our three research questions.

### A. Single-Tile Baseline (RQ1)

Single-tile experiments established baseline characteristics. Integer achieved 4.25 MCUPS while floating-point achieved 5.5 MCUPS (single-thread) and 7.5 MCUPS (six-thread)—24% degradation for integer confirming IPU's floating-point optimization. The 1.4× speedup from six-thread parallelization demonstrates thread-level parallelism importance. Maximum sequence length reached 780 elements (600 KB of 624 KB utilized). Extension to 1,472 tiles showed near-linear scaling: 10.75 GCUPS (six-thread) (Figure 1).

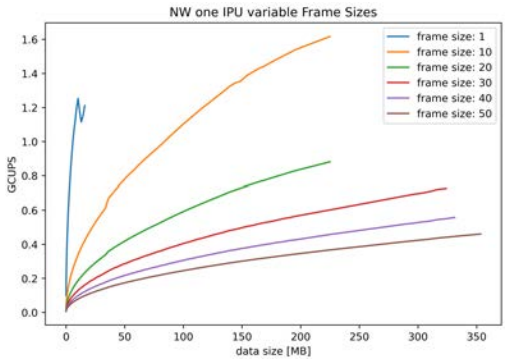


Fig. 2: Frame size impact: smaller frames achieve higher throughput but face memory limits; larger frames enable 4-22× larger sequences at reduced performance.

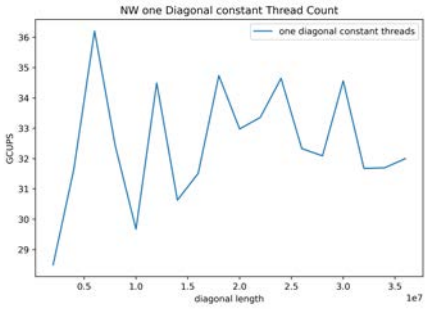


Fig. 3: Optimal diagonal computation: 29-36 GCUPS for uniform workloads, scaling from 550k to 36M elements.

### B. Frame Size Impact (RQ2)

Frame=1 achieved 1.25 GCUPS peak for 1,800-length sequences but limited to 2,000-length maximum. Frame=10 achieved 1.6 GCUPS at 3,500-length. Frame=50 reduced performance to 0.25 GCUPS but extended capacity to 9,400-length—22× larger sequences.

Memory profiling: Frame=1 required 137 MB control code + 236 MB vertex instances vs 27 MB data—13.8× overhead. Frame=50: 10 MB control + 8 MB vertex instances but 393 MB exchange code, demonstrating finer granularity incurs vertex overhead while coarser shifts to communication (Figure 2).

### C. Optimal Diagonal Performance (RQ1)

Single diagonal benchmarks with dynamic distribution maintaining 8,832 threads achieved 29-36 GCUPS for 550k to 36M elements, exceeding GPU baseline (15 GCUPS [4]) by 2.4×. Naive approach achieved 15-30 GCUPS, limited to 550k elements. Dynamic distribution extended capacity 65× while maintaining peak performance (Figure 3).

### D. Three-Phase Implementation (RQ 1 & 2)

Complete three-phase implementation averaged 3-7 GCUPS with maximum 9,000 nucleotides. Build-up and culmination

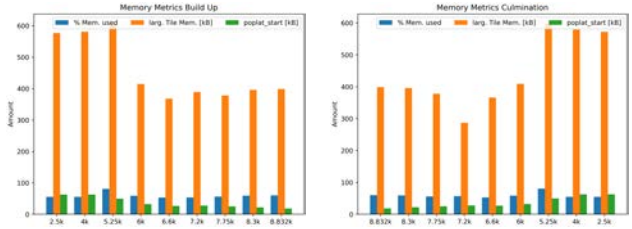


Fig. 4: Memory metrics for build-up (left) and culmination (right) phases across diagonal intervals. Bars show % memory used, largest tile memory (kB), and `poplar_start` size (kB).

TABLE I: Performance Comparison: IPU versus GPU

| Platform/Config | GCUPS | Max Length | Power |
|---|---|---|---|
| *IPU (GC200)* | | | |
| Single-tile (6 threads) | 0.0075 | 780 | 150W |
| Frame=1 (fine) | 1.25 | 2,000 | 150W |
| Frame=50 (coarse) | 0.25 | 9,400 | 150W |
| Optimal diagonal | 36 | 36M (diag) | 150W |
| Three-phase | 3-7 | ∼9,000 | 150W |
| *GPU (A30)* | | | |
| Barrier sync [2] | 0.2 | 30,000 | 165W |
| Coop. groups [4] | 5-15 | 55,000 | 165W |

showed 1.25-3.0 GCUPS due to memory complexity, requiring nine program launches with `poplar_start` growing to 63 KB. Figure 4 shows memory usage across diagonal intervals for both phases: peak tile memory reaches 600 KB (of 624 KB) at high-complexity intervals, while `poplar_start` varies from 18–63 KB. Saturated phase achieved 6.5 GCUPS—1/6 of optimal—demonstrating memory overhead impact.

### E. Performance Comparison

Table I summarizes implementations. IPU achieves 2.4× better performance for uniform workloads but faces severe constraints: 9,400 vs 55,000 nucleotides. Realistic IPU performance (3-7 GCUPS) is comparable to GPU (5-15 GCUPS) but with 6× lower capacity.

## IV. Discussion

### A. Performance Analysis (RQ1)

IPU achieves competitive performance for uniform workloads (36 vs 15 GCUPS, 2.4×) with superior energy efficiency (0.24 vs 0.09 GCUPS/Watt, 2.7×). However, realistic implementations show comparable performance (3-7 vs 5-15 GCUPS) with lower capacity. The gap between optimal (36) and realistic (3-7 GCUPS) reveals memory overhead dominates.

### B. Architectural Constraints (RQ2)

Despite 62.5 TFlops/s theoretical performance, 13.8× memory overhead from compiler-generated code severely underutilizes capacity. These memory challenges are consistent with our prior observations in butterfly factorization work for vision transformers [6], where memory reduction techniques proved essential for scaling to multi-IPU systems. Fine-grained parallelism (frame=1) maximizes performance but limits to 2,000 elements; coarse-grained (frame=50) enables 22× larger sequences at 75% performance reduction. GPU's larger memory (24 GB vs 918 MB) provides 6× sequence length advantage.

### C. Bioinformatics Applications (RQ3)

**Suitable:** (1) Database searches—batch processing short sequences (100-1,000 nt); (2) NGS quality control—Illumina reads (50-300 nt) with frame=1; (3) Population genetics—batch processing 1,000+ individuals.

**Unsuitable:** (1) Reference genome alignment—chromosomes (50-260M nt) exceed memory; (2) Long-read alignment—PacBio/Nanopore (10k-100k+ nt) require frame=50 with 75% penalty; (3) De novo assembly.

**Guidance:** Use frame=1 for short reads (150-300 nt) achieving 1.25 GCUPS; frame=50 for longer sequences. Use floating-point to avoid 24% integer penalty. IPU's 2.7× energy efficiency benefits high-throughput centers.

### D. Related Work

Our work complements prior IPU bioinformatics research [1], [7] on many-small-sequences (X-Drop, batch alignment). We provide single-large-sequence characterization, revealing memory constraints absent in batch applications. Our optimal case (36 GCUPS) exceeds GPU implementations [2], [4], validating computational potential.

## V. Conclusion

IPU achieves 2.4× performance advantage and 2.7× energy efficiency over GPU for uniform workloads, but memory limits restrict applications to sequences under 9,400 vs 55,000 nucleotides. IPUs complement rather than replace GPUs—excelling at batch-parallel operations on short sequences. Key findings: (1) 13.8× memory overhead dominates, (2) Frame sizing enables 22× capacity at 75% performance cost, (3) Floating-point provides 24% advantage. Future work should explore multi-IPU scaling [6] and hybrid architectures.

## References

[1] L. Burchard et al., "Space efficient sequence alignment for sram-based computing: X-drop on the graphcore ipu," in SC Conference 2023.
[2] A. Chaudhary et al., "A gpu based implementation of needleman-wunsch algorithm using skewing transformation," in *8th International Conference on Contemporary Computing*, pp. 498–502.
[3] Z. Jia et al., "Dissecting the graphcore ipu architecture via microbenchmarking," 2019. [Online]. Available: https://arxiv.org/abs/1912.03413
[4] F. Kurt et al., "Implementations of the needleman-wunsch algorithm for gpu architectures," in *7th National High Performance Computing Conference*, 2022, Istanbul, Turkey.
[5] S. B. Needleman and C. D. Wunsch, "A general method applicable to the search for similarities in the amino acid sequence of two proteins," *Journal of Molecular Biology*, vol. 48, no. 3, pp. 443–453, 1970.
[6] S.-K. Shekofteh et al., "Butterfly factorization for vision transformers on multi-ipu systems," *Parallel Computing*, vol. 127, p. 103165, 2026.
[7] M. Zhao et al., "ipuma: High-performance sequence alignment on the graphcore ipu," in *39th ISC High Performance 2024 Research Paper Proceedings*, 2024, pp. 1–11.

 DOI 10.15203/99106-194-6-22

# Comparison of *UQLab* and *pygpc* for fast approximation of leadfield matrices in EEG/MEG source analysis

M. Schmatz, D. Baumgarten, and J. Vorwerk

***Abstract*—Accurate quantification of uncertainty in EEG/MEG source analysis requires the computation of numerous leadfield matrices for varying conductivity parameters in the head volume conductor model. Since each leadfield computation is numerically expensive, surrogate models based on generalized Polynomial Chaos (gPC) expansions are employed to decrease the runtime. These surrogates must approximate leadfield matrices with respect to varying conductivities as accurately as possible while keeping computational costs low. In this study, we compare two open-source libraries for gPC surrogate construction, *UQLab* and *pygpc*. Using identical input data in the form of precomputed leadfield matrices, we evaluate their performance in terms of computational efficiency and approximation accuracy. Our results show that, for comparable error levels, *pygpc* achieves substantially faster runtimes than *UQLab*, demonstrating its suitability for large-scale uncertainty quantification in EEG/MEG forward modelling.**

***Index Terms*—conductivity uncertainty, EEG/MEG source analysis, generalized polynomial chaos, uncertainty quantification**

## I. Introduction

Electroencephalography (EEG) and magnetoencephalography (MEG) source analysis has become an important tool for the localization of brain activity with a high temporal resolution [1]. Source analysis has several applications in research and clinical practice, for example the localization of interictal epileptic discharges in presurgical epilepsy diagnosis (PED). Therefore, achieving a high spatial accuracy in EEG/MEG source analysis is desirable. An important step to achieve this is the use of individual head volume conductor models, which accurately discretize the individual head geometry. Nowadays, head models represent several head tissues with differing conductivities, such as white matter, gray matter, cerebrospinal fluid, skull, or skin. Whereas the individual geometry of the head can easily be determined from magnetic resonance images (MRIs), the conductivities can only be determined from the literature, neglecting inter-individual variances. Therefore, tissue conductivity variations are a source of significant uncertainties particularly in EEG source analysis [2, 3].

Uncertainty quantification in EEG/MEG source analysis seeks to address precisely this issue, as it attempts to quantify and visualize such uncertainty as well as to reduce it. However, performing uncertainty quantification – whether for combined EEG/MEG or for single modality EEG or MEG data – requires computationally intensive calculations.

M. Schmatz, D. Baumgarten, and J. Vorwerk are with the Biomedical Engineering Group, Department of Mechatronics, University of Innsbruck, Innsbruck, Austria. E-mail: Mathias.Schmatz@uibk.ac.at

EEG/MEG source analysis consists of two steps: the forward problem and the inverse problem. The EEG/MEG forward problem is the simulation of the electric or magnetic field generated by a dipolar current source inside the brain. The accuracy of an EEG/MEG forward solution depends on the chosen numerical method and the accuracy of the head volume conductor. Commonly, forward solutions for several source positions, the so-called source space, are concatenated to yield a single matrix $L$, referred to as the leadfield or gain matrix. This matrix provides a (simulated) mapping of source activity inside the head to the measurement sensors and can be used as basis to solve the inverse problem.

Given a measurement, the inverse problem is the task of finding the active brain area causing the measured signal. In contrast to the EEG/MEG forward problem, which was shown to admit a unique solution, the inverse problem is ill-posed as an infinite amount of solutions exists [4]. To ensure uniqueness, prior information about the behavior of the solution has to be taken into account. Therefore, numerous inverse methods associated to different prior assumptions exist, resulting in differently accurate inverse solutions for a certain distribution of brain activity. Due to its reliance on the leadfield matrix $L$, the accuracy of the inverse problem crucially depends on the solution of the forward problem [5].

Therefore, it is desirable to quantify and visualize how the variability of uncertain parameters, such as the tissue conductivities, affects the estimated neural activity. Such an uncertainty quantification can, for example, be obtained through Monte Carlo techniques [6], which require the computation of numerous leadfield matrices for different conductivity distributions. Since computing each leadfield matrix is time-consuming (approximately 40 minutes per conductivity distribution), exhaustive evaluations are not directly feasible in practical applications. This is particularly limiting because Monte Carlo methods converge only at a rate of $\mathcal{O}(N^{-1/2})$, where $N$ denotes the number of samples.

To circumvent this problem, the leadfield matrices can be approximated using *generalized polynomial chaos expansions* (gPC) [7], which allow the rapid approximation of numerous leadfield matrices in a reasonable time, as only a reduced number of leadfield matrices has to be calculated. Once a gPC expansion is initialized, the approximations rely only on the evaluation of multivariate polynomials.

A gPC expansion represents a mapping $L(\sigma)$, which, in our

 DOI 10.15203/99106-194-6-22

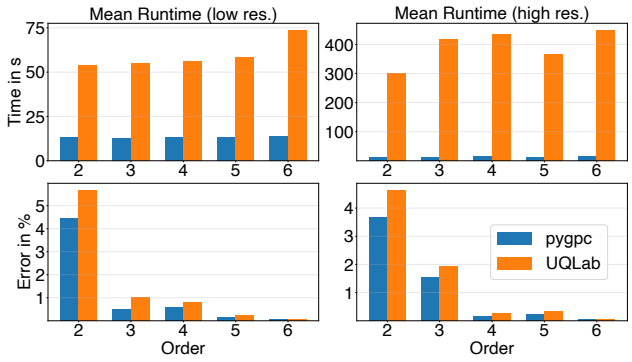


Fig. 1: Total runtime and mean RMSE error for approximating ten random conductivity combinations for 66 sources (left) and 132 sources (right) and different gPC orders.

TABLE I: Conductivity intervals for brain, skull, and skin tissues in mS/m.

| Tissue | Min | Max. | Standard | Reference |
|---|---|---|---|---|
| Brain | 220 | 670 | 330.0 | Haueisen et al., 2002 |
| Skull | 1.6 | 33.0 | 10.0 | Akhtari et al., 2002; Dannhauer et al., 2011; Hoekema et al., 2003 |
| Skin | 280 | 870 | 430.0 | Haueisen et al., 2002 |

case, assigns to each conductivity vector $\sigma = (\sigma_1, \ldots, \sigma_i)$ its corresponding leadfield matrix, as the series $L(\sigma) = \sum_{\alpha \in \mathbb{N}^i} y_\alpha \, \psi_\alpha(\sigma)$. The basis functions $\psi_\alpha$ are multivariate polynomials that are orthonormal with respect to the joint probability distribution of $\sigma$, i.e., $\langle \psi_\alpha, \psi_\beta \rangle = \mathbb{E}[\psi_\alpha \, \psi_\beta] = \delta_{\alpha\beta}$, where $\delta_{\alpha\beta}$ denotes the Kronecker delta, which equals 1 if $\alpha = \beta$ and 0 else.

In practice, the infinite series $\sum_{\alpha \in \mathbb{N}^i} y_\alpha \, \psi_\alpha(\sigma)$ is truncated after a finite number of terms. Usually, one only includes multi-indices that satisfy the condition $\|\alpha\|_\infty \leq p$ for some natural number $p$, which is referred to as the order of the gPC expansion, although other truncation schemes are also possible.

The gPC coefficients $y_\alpha$ can then be obtained either by regression-based approaches, such as ordinary least squares regression or Tikhonov regularization, or by quadrature-based approaches that exploit the orthogonality of the polynomial basis. In both of these cases, the function $L$ has to be evaluated at a suitable set of predefined grid points. The choice of the grid points constitutes an additional degree of freedom of the method. The number of grid points is not to be confused with the order $p$ of the gPC expansion, which determines the number of basis functions included in the approximation.

In this study, we compare the performance of two open-source toolboxes for gPC computations in the approximation of EEG leadfield matrices with regard to computation time and accuracy.

## II. Methods

We used the open-source library MNE-Python to generate a Boundary Element Method (BEM) volume conductor model for one of the predefined head models provided by the software. The model distinguishes three tissue layers: skin, skull and brain. For various conductivity combinations $\sigma = (\sigma_{\text{brain}}, \sigma_{\text{skull}}, \sigma_{\text{skin}})$, we calculated the forward solutions for several source positions, that is, the associated leadfield matrix $L(\sigma)$. To reduce computation time, we used leadfields of a reduced number of sources of 66 and 132 in our study. The number of electrodes was 59.

We chose to use a quadrature-based approach for the computation of the gPC coefficients, as it easily allows us to use identical input data in both libraries. For this calculation, we assumed the tissue conductivities to be Beta distributed with shape parameters $\alpha = \beta = 3$ and support intervals given in Table I. This choice is motivated by the fact that head tissue conductivities are naturally bounded and are expected to exhibit a distribution with a bell-shaped probability distribution, see [12] or [13].

For the evaluation of $L$, we employed a Jacobi-type grid consisting of $p + 1$ grid points in each dimension, yielding a total of $(p+1)^3$ grid points, where $p$ denotes the polynomial order of the gPC expansion. Based on precomputed leadfields for the respective grids, we constructed gPC expansions of various polynomial orders with both *pygpc* and *UQLab*. In addition to the requirement $\|\alpha\|_\infty \leq p$, we imposed the commonly used constraint $\|\alpha\|_1 \leq p$ to further reduce computation time, as is standard in high-dimensional uncertainty quantification. Note that the second condition is more restrictive than the first and therefore implicitly includes it. We tested expansion orders ranging from 2 to 6 to assess the accuracy.

For ten random sets of conductivities, we calculated exact solutions for the leadfield matrices $L(\sigma_i)$ to serve as "ground truth" to compare the approximations of *UQLab* and *pygpc* against. For each conductivity configuration, we approximated the associated leadfield matrix using a gPC expansion in both libraries using a quadrature based approach and computed the root mean square error (RMSE) for each approximation of the matrices $L(\sigma_i)$, as well as the percentage error, i.e. the norm of the difference normalized by the norm of the exact solution. In addition, we evaluated the runtime of the gPC computations for both libraries.

All computations were performed on a workstation with an AMD Ryzen AI 7 PRO 350 CPU (8 cores, 16 threads) and 32 GB RAM.

## III. Results

The computation times and accuracies for the leadfields with 66 and 132 source positions are shown in Figure 1. We clearly observe that, for comparable levels of error, *pygpc* achieves substantially lower computation times than *UQLab*. In most cases, *pygpc* also attains slightly lower approximation errors. Only for relatively small dimensions (e.g. $3 \times 59$) *UQLab* outperforms *pygpc* in terms of both error and computation time (not depicted here).

However, as dimensions increase, the computation times of *UQLab* significantly increase in comparison to *pygpc*. For a gPC approximation of order 4 and a leadfield matrix of dimensions $59 \times 258$, *UQLab* already required approximately 70

minutes, whereas *pygpc* completed the computation in around 12 seconds, both achieving an approximation error below $1\%$ (not depicted here). At polynomial orders between two and four, the approximation error decreases rapidly, yielding substantial improvements in accuracy. With further increasing order, the convergence rate gradually slows down and further gains in accuracy become marginal.

## IV. Discussion

Our experiments show that, for approximating leadfield matrices dependent on the tissue conductivities $\sigma$, the Python library *pygpc* requires substantially less computation time than its MATLAB counterpart *UQLab*, while even achieving slightly lower errors. Our results confirm previous observations reported in, for example, [7] and [14], suggesting that a polynomial order of $4$ suffices to obtain satisfactory accuracy. Increasing the polynomial order further only has negligible improvements in approximation accuracy. Although computation time does not increase significantly beyond order 4, the additional computational effort does not result in significant accuracy gains.

To simplify our calculations, we restricted our evaluations to rather small leadfield matrices in this study. Whereas we only evaluated leadfield matrices with maximally 258 columns, several thousand columns are common in practice. As expected, our findings suggest that increasing the dimensionality of the leadfield matrices has little effect on the approximation accuracy. However, the computational cost increases significantly with the matrix dimensions when using *UQLab*, whereas *pygpc* shows only moderate increases, making it more suitable for large-scale problems. Besides the number of sources, also the number of sensors might be larger in practice, which is expected to have comparable effects.

As a further restriction, the present study focuses on a limited set of only three parameters. The approach naturally extends to larger parameter spaces, allowing the inclusion of additional tissue conductivities. State-of-the-art highly-realistic individual head volume conductor models commonly distinguish five or more conductive compartments. Increasing the number of distinguished parameters would of course increase the complexity of the gPC expansions, requiring both the computation of a larger amount of leadfield matrices and a more complex calculation of the gPC expansions itself.

In future work, the influence of higher dimensions of leadfield matrices and parameter space should be evaluated. Our experiments further indicate that the approximation error exhibits a spatial structure, suggesting that certain sensors are systematically overestimated while others are underestimated, rather than the error being uniformly distributed across the leadfield matrix (not depicted here). A more detailed analysis of this behaviour is therefore desirable. Furthermore, comparisons of different algorithms for computing the gPC coefficients, such as regression-based methods and adaptive regression approaches, should be performed. In addition, parallelization strategies could be exploited to further reduce computation times.

## Acknowledgment

This research was funded in whole or in part by the Austrian Science Fund (FWF) [10.55776/P35949, 10.55776/PIN3777624]. For open access purposes, the author has applied a CC BY public copyright licence to any author accepted manuscript version arising from this submission.

AI-based language tools were used to support the formulation and linguistic refinement of the manuscript.

## References

[1] R. Brette and A. Destexhe, *Handbook of neural activity measurement*. Cambridge University Press, 2012.

[2] J. Vorwerk, Ü. Aydin, C. H. Wolters, and C. R. Butson, "Influence of head tissue conductivity uncertainties on EEG dipole reconstruction," *Frontiers in neuroscience*, vol. 13, p. 531, 2019.

[3] J. Vorwerk, C. H. Wolters, and D. Baumgarten, "Global sensitivity of EEG source analysis to tissue conductivity uncertainties," *Frontiers in Human Neuroscience*, vol. 18, p. 1335212, 2024.

[4] J. Sarvas, "Basic mathematical and electromagnetic concepts of the biomagnetic inverse problem," *Physics in Medicine & Biology*, vol. 32, no. 1, p. 11, 1987.

[5] J.-H. Cho, J. Vorwerk, C. H. Wolters, and T. R. Knösche, "Influence of the head model on EEG and meg source connectivity analyses," *Neuroimage*, vol. 110, pp. 60–77, 2015.

[6] C. Schmidt, P. Grant, M. Lowery, and U. van Rienen, "Influence of uncertainties in the material properties of brain tissue on the probabilistic volume of tissue activated," *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 5, pp. 1378–1387, 2012.

[7] C. Schmidt, S. Wagner, M. Burger, U. v. Rienen, and C. H. Wolters, "Impact of uncertain head tissue conductivity in the optimization of transcranial direct current stimulation for an auditory target," *Journal of Neural Engineering*, vol. 12, no. 4, p. 046028, jul 2015. [Online]. Available: https://doi.org/10.1088/1741-2560/12/4/046028

[8] J. Haueisen, C. Ramon, M. Eiselt, H. Brauer, and H. Nowak, "Influence of tissue resistivities on neuromagnetic fields and electric potentials studied with a finite element model of the head," *IEEE Transactions on Biomedical Engineering*, vol. 44, no. 8, pp. 727–735, 2002.

[9] M. Akhtari, H. Bryant, A. Mamelak, E. Flynn, L. Heller, J. Shih, M. Mandelkem, A. Matlachov, D. Ranken, E. Best *et al.*, "Conductivities of three-layer live human skull," *Brain topography*, vol. 14, no. 3, pp. 151–167, 2002.

[10] M. Dannhauer, B. Lanfer, C. H. Wolters, and T. R. Knösche, "Modeling of the human skull in EEG source analysis," *Human brain mapping*, vol. 32, no. 9, pp. 1383–1399, 2011.

[11] R. Hoekema, G. Wieneke, F. Leijten, C. Van Veelen, P. Van Rijen, G. Huiskamp, J. Ansems, and A. Van Huffelen, "Measurement of the conductivity of skull, temporarily removed during epilepsy surgery," *Brain topography*, vol. 16, no. 1, pp. 29–38, 2003.

[12] K. Weise, L. Di Rienzo, H. Brauer, J. Haueisen, and H. Toepfer, "Uncertainty analysis in transcranial magnetic stimulation using nonintrusive polynomial chaos expansion," *IEEE Transactions on Magnetics*, vol. 51, no. 7, pp. 1–8, 2015.

[13] G. B. Saturnino, A. Thielscher, K. H. Madsen, T. R. Knösche, and K. Weise, "A principled approach to conductivity uncertainty analysis in electric field calculations," *Neuroimage*, vol. 188, pp. 821–834, 2019.

[14] B. Sudret, M. Berveiller, and M. Lemaire, "A stochastic finite element procedure for moment and reliability analysis," *European Journal of Computational Mechanics/Revue Européenne de Mécanique Numérique*, vol. 15, no. 7-8, pp. 825–866, 2006.

 DOI 10.15203/99106-194-6-23

# Predicting CPAP Adherence with Machine Learning

T.F.A. Steinbrinker, M.C. Maurer, P. Zaschke, D. Waldmannstetter, M. Arzt, A.C. Hauschild, D. Krefting,

***Abstract*—Poor adherence to Continuous Positive Airway Pressure (CPAP) therapy limits the effectiveness of treatment for Obstructive Sleep Apnea (OSA). Early identification of patients at risk of non-adherence could enable targeted interventions. Using data from the Apnea Positive Pressure Long-term Efficacy Study (APPLES), we evaluated the ability of explainable machine learning models to predict CPAP adherence after six months. A total of $689$ patients were analyzed using demographic, clinical, and early adherence features. Six machine learning models were trained and compared. All models showed similar performance, with the multilayer perceptron achieving the highest accuracy ($0.79$) and AUC ($0.85$). Feature importance analysis highlighted early CPAP usage as the strongest predictor of long-term adherence. These results demonstrate the potential of machine learning to support early identification of patients at risk of CPAP non-adherence.**

***Index Terms*—obstructive sleep apnea, CPAP adherence, machine learning**

## I. Introduction

Obstructive Sleep Apnea (OSA) is a highly prevalent sleep disorder affecting millions worldwide, significantly impacting quality of life and health outcomes [1], [2]. Characterized by partial or complete obstruction of the upper airway during sleep, OSA leads to disrupted sleep patterns and associated health issues such as cardiovascular problems and cognitive impairment. Continuous Positive Airway Pressure (CPAP) therapy is the gold standard treatment, effectively maintaining upper airway patency and improving sleep quality. The efficacy of CPAP is dose-dependent; higher usage hours correlate with better clinical outcomes, underscoring the importance of adherence. However, many patients face challenges in maintaining consistent CPAP usage, noncompliance (defined as use for $< 4$ h/night) have been reported to range from $29$ to $83\%$ in OSA patients receiving long-term therapy [3]. Identifying and supporting patients at risk of low adherence is crucial for enhancing treatment effectiveness. The rise of big datasets [4], [5] considering CPAP adherence has given a chance to analyze termination rates and investigate factors predictive of therapy termination. Most studies only investigate a single or a few factors for CPAP adherence. Meanwhile, machine learning (ML)-based data analysis enables us to uncover hidden patterns, including various features in sleep medicine [6].

In this paper, we identify a set of predictive features, which served as the basis for training multiple ML models. We show preliminary model results and analyze feature importance using an explainable AI–based analysis. The intended clinical use case of the proposed models is early risk stratification during the first weeks of CPAP therapy. Specifically, we aim to predict six-month adherence based on baseline characteristics and early usage patterns in order to identify patients who may benefit from intensified follow-up or telemonitoring interventions.

TS, MM, PZ, DF are with the Department of Medical Informatics, University Medical Center Göttingen, Germany, e-mail: tabea.steinbrinker@med.uni-goettingen.de.

DW, MA are with the Universityclinic Regensburg, Germany.

ACH is with the Justus-Liebig-University, Gießen, Germany.

## II. Methods

### A. Dataset

We used the Apnea Positive Pressure Long-term Efficacy Study (APPLES) [4], [5]; a NHLBI-sponsored 6-month, multicenter trial conducted at $5$ U.S. university hospitals, or private practices. It consists of $1,516$ patients, enrolled from November 2003, each followed up for up to $6$ months over $11$ visits. $1,105$ patients received active vs. sham CPAP devices. The NSRR APPLES includes, among other variables, data on daytime sleepiness, the maintenance of wakefulness test, mood, quality of life, PAP adherence, vitals weight, demographic and other baseline characteristics.

### B. Feature selection

A set of features was chosen to predict CPAP adherence after $6$ months, defined as the usage of $> 4$h/night for more than $70\%$ of the nights within a month. The features, together with their mean and standard deviation or their distribution is shown in Table I. The set of features is based on a sleep expert's assesment of their relevance for adherence prediction. To ensure more balanced features, the cardiac diseases myocardial infarction (MI), heart failure (HF), coronary artery disease (cad), and stroke have been merged into one feature, labeled $1$, if any of the diseases is present. Patients, where one of the features was not given, where excluded from the dataset, resulting in $689$ patients. This complete-case approach may introduce selection bias and should be considered when interpreting results.

### C. Machine Learning

We compare six different ML methods that are available in the Python library scikit-learn: Logistic Regression, Support Vector Machine (SVM), a small Multilayer Perceptron (MLP), Decision Tree, Random Forest and XGBoost. Unlike deep learning models, they offer explanations out of the box and are therefore preferred in preliminary feature identification [7]. For all models, the dataset is split into a train- and a test set with a $80\%/20\%$ split. In the case of Logistic Regression, SVM, and MLP, the data is standardized by mean and standard deviation to improve training. To account for class imbalance, Logistic Regression, SVM, Decision Tree, Random Forest and XGBoost were trained using class-weight balancing. The MLP classifier did not directly support class weighting and was

TABLE I: Descriptive Analysis for the Used Features.

| Feature | Unit | Value |
|---|---|---|
| Number of patients | | 945 |
| TX Group "Active" | % | 52.28 (N=364) |
| Male | % | 66.2 (N=625) |
| Age | Years | $51.57 \pm 12.17$ |
| Height | cm | $172.91 \pm 10.35$ |
| Weight | kg | $96.95 \pm 22.97$ |
| BMI | kg/m² | $32.13 \pm 7.15$ |
| Smoker | % | 31.35 (N=216) |
| Diabetes | % | 7.08 (N=67) |
| Hypertension | % | 31.96 (N=302) |
| CAD, MI, HF or Stroke | % | 2.90 (N=20) |
| Asthma | % | 11.64 (N=110) |
| Snoring | Years | $17.39 \pm 12.39$ |
| Depression | % | 23.07 (N=218) |
| ESSQ at Baseline | Points | $10.43 \pm 4.42$ |
| ESSQ at CPAP Titration | Points | $9.68 \pm 4.39$ |
| 2-month Adherence | Hours | $4.481 \pm 2.319$ |
| 6-month Adherence | % | 31.06 (N=214) |

TABLE II: Descriptive Analysis for the used Features.

| Model | Accuracy | AUC | Precision | Recall | F1 |
|---|---|---|---|---|---|
| MLP | **0.79** | **0.85** | 0.72 | 0.67 | 0.69 |
| XGBoost | 0.72 | 0.85 | **0.92** | 0.22 | 0.36 |
| SVM | 0.76 | 0.82 | 0.62 | 0.84 | **0.71** |
| Decision Tree | 0.71 | 0.83 | 0.56 | **0.86** | 0.68 |
| Random Forest | 0.78 | 0.82 | 0.67 | 0.76 | 0.71 |
| Log Regression | 0.76 | 0.84 | 0.63 | 0.80 | 0.70 |

therefore trained without explicit rebalancing. Then, a grid search is implemented for hyperparameter tuning using a 5-fold cross-validation. All models are evaluated on the test set. The code, trained models, and chosen parameters can be found in the attached repository. The models are evaluated based on the different metrics Accuracy, Area Under Curve (AUC), Precision, Recall, and F1 Score.

## III. Results

Table II summarizes the performance of the trained ML models. Highest values for the specific metrics are marked in bold. Note that as non-adherence is the prediction requiring intervention, precision and recall are computed with non-adherence as the positive outcome. Figure 1 shows the receiver operating characteristic (ROC) curve of the different models, where again, the positive value is non-adherence. We note that all the models have comparable performances, with MLP yielding the highest AUC and Accuracy. Figure 2 shows the feature importance of the XGBoost Model, as it, a tree-based ensemble model, provides stable and well-established feature attribution measures. Although its F1 score was lower due to a precision–recall trade-off, its AUC was among the highest, and tree-based importance metrics are more robust than coefficient-based or kernel-based alternatives. We clearly see the high importance of the previous adherence, and the similarity in

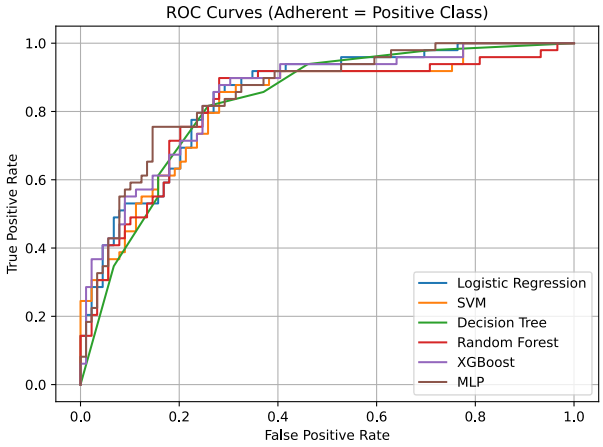


Fig. 1: Receiver Operating Characteristic Curve.

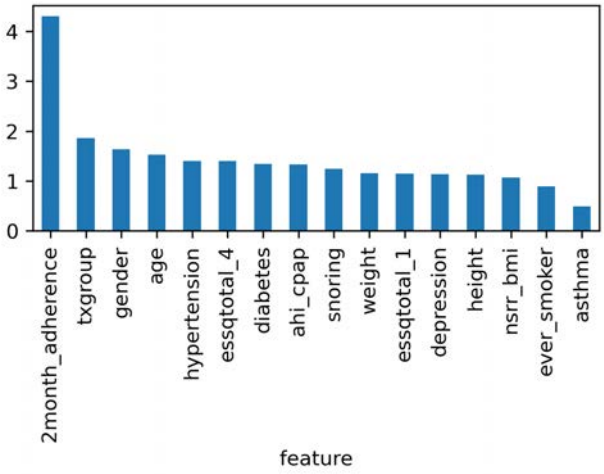


Fig. 2: Feature Importance for the XGBoost Model.

the feature importance of the demographic data. Furthermore, 'txgroup' has a high feature importance, indicating whether a patient received CPAP therapy or a sham device, indicating the need for further investigation of the causal relationship between device type and adherence.

## IV. Discussion

With these models, we are able to predict the outcome of CPAP adherence after 6 months, based on different demographic features and the adherence after 2 months. Our models yielded an AUC of up to 0.85. By choosing the right threshold, we can manage to correctly identify the non-adherent class, giving room for intervention methods, such as further telemonitoring guidance [8]. The use of this dataset, a U.S.-based cohort, enables reproducible benchmarking of ML approaches for CPAP adherence prediction. Validation on German or other national cohorts would be an important next step to assess generalizability across healthcare systems and device infrastructures and is planned for future work.

## V. Limitations

Due to restricted data availability, this study has several limitations. The complete-case feature selection strategy may introduce selection bias and reduce representativeness. The APPLES dataset is historical and lacks high-resolution telemetry data from modern CPAP devices, limiting feature richness. Adherence was modeled as a binary outcome based on a clinical threshold rather than as a continuous measure, and feature importance results should not be interpreted causally, particularly regarding treatment group effects. Although class imbalance was addressed through model weighting, and multiple models showed consistent behavior, no external validation cohort was available; thus, results should be interpreted as methodological benchmarking rather than deployment-ready performance.

## VI. Conclusion

This study demonstrates that machine learning models can effectively predict long-term CPAP adherence in patients with obstructive sleep apnea using routinely collected clinical and early usage data. Across multiple explainable models, we achieved an AUC of up to 0.85, with early CPAP adherence emerging as the strongest predictor of six-month outcomes. These findings highlight the value of short-term adherence monitoring for identifying patients at risk of non-adherence. By enabling early and targeted interventions, such as personalized coaching or telemonitoring support, ML-based adherence prediction has the potential to improve long-term therapy compliance and overall treatment effectiveness.

## VII. Code Availability

The code is available at: https://gitlab.gwdg.de/medinfpub/kvf-group. Please fully respect the usage licenses of the underlying components.

## Acknowledgment

This work was partially funded by the German Federal Ministry of Research, Technology and Space, Grant numbers 01ZZ2324A, 01ZZ2324C, (Somnolink).

## References

[1] A. V. Benjafield, N. T. Ayas, P. R. Eastwood, R. Heinzer, M. S. M. Ip, M. J. Morrell, C. M. Nunez, S. R. Patel, T. Penzel, J.-L. Pépin, P. E. Peppard, S. Sinha, S. Tufik, K. Valentine, and A. Malhotra, "Estimation of the global prevalence and burden of obstructive sleep apnoea: a literature-based analysis," *The Lancet Respiratory Medicine*, vol. 7, pp. 687–698, Aug. 2019.

[2] P. Lévy, M. Kohler, W. T. McNicholas, F. Barbé, R. D. McEvoy, V. K. Somers, L. Lavie, and J.-L. Pépin, "Obstructive sleep apnoea syndrome," *Nature Reviews Disease Primers*, vol. 1, p. 15015, June 2015. Publisher: Nature Publishing Group.

[3] H. Woehrle, M. Arzt, A. Graml, I. Fietze, P. Young, H. Teschler, and J. H. Ficker, "Predictors of positive airway pressure therapy termination in the first year: Analysis of big data from a german homecare provider," *BMC Pulmonary Medicine*, vol. 18, Dec 2018.

[4] S. F. Quan, C. S. Chan, W. C. Dement, A. Gevins, J. L. Goodwin, D. J. Gottlieb, S. Green, C. Guilleminault, M. Hirshkowitz, P. R. Hyde, G. G. Kay, E. B. Leary, D. A. Nichols, P. K. Schweitzer, R. D. Simon, J. K. Walsh, and C. A. Kushida, "The association between obstructive sleep apnea and neurocognitive performance–the Apnea Positive Pressure Long-term Efficacy Study (APPLES)," *Sleep*, vol. 34, pp. 303–314B, Mar. 2011.

[5] G.-Q. Zhang, L. Cui, R. Mueller, S. Tao, M. Kim, M. Rueschman, S. Mariani, D. Mobley, and S. Redline, "The National Sleep Research Resource: towards a sleep data commons," *Journal of the American Medical Informatics Association: JAMIA*, vol. 25, pp. 1351–1358, Oct. 2018.

[6] D. Krefting, M. Arzt, M. Brandt, M. Goldammer, F. Ehrlich, J. T. Maurer, T. Penzel, T. Sehr, A. Rodenbeck, P. Zaschke, and C. Schöbel, "Somnolink – a national initiative to improve data sharing for better healthcare and research in obstructive sleep apnea," *GMS Med Inform Biom Epidemiol*, vol. 21, Sept. 2025. Archive Location: ddc:Medizin, Gesundheit Publisher: German Medical Science GMS Publishing House;Düsseldorf.

[7] M. C. Maurer, J. M. Metsch, P. Hempel, T. Bender, N. Spicher, and A.-C. Hauschild, "Explainable Artificial Intelligence on Biosignals for Clinical Decision Support," in *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, KDD '24, (New York, NY, USA), pp. 6597–6604, Association for Computing Machinery, Aug. 2024.

[8] H. Woehrle, M. Arzt, A. Graml, I. Fietze, P. Young, H. Teschler, and J. H. Ficker, "Effect of a patient engagement tool on positive airway pressure adherence: Analysis of a german healthcare provider database," *Sleep Medicine*, vol. 41, p. 20–26, Jan 2018.

 DOI 10.15203/99106-194-6-24

# WavePrep - Create Machine-Learning-Ready Datasets from MIMIC Waveforms

M. Elwes, O. Beyan, E. Kutafina

***Abstract*—The Medical Information Mart for Intensive Care (MIMIC) waveform databases provide high-resolution physiological signals from intensive care units but remain underutilised compared to the relational MIMIC core databases. One barrier is the lack of accessible tools for transforming waveform data stored in the Waveform Database (WFDB) format into machine-learning-ready datasets. This work introduces WavePrep, an open-source software framework that addresses this gap by enabling flexible, transparent, and efficient preprocessing of MIMIC waveform data. WavePrep allows users to define fine-grained, channel-specific preprocessing pipelines via a user-friendly JSON configuration, supporting common signal processing steps such as downsampling, data cleaning, imputation, windowing, and dataset splitting. The framework automatically detects dependencies between preprocessing steps, prevents data leakage through patient-level splitting, and exploits parallelisation to accelerate dataset creation. We evaluate WavePrep by creating multiple datasets derived from the MIMIC-III matched waveform database, including numeric and non-numeric signals. Experimental results on a multi-core server demonstrate robust and scalable performance, with processing times between 33.10 and 50.53 milliseconds per hour of recording. Datasets created using WavePrep were successfully applied in machine learning tasks. By lowering technical barriers and providing transparent preprocessing metadata, WavePrep facilitates broader adoption of clinical waveform data and improves the AI-readiness of signal datasets. GitHub: https://github.com/BI-K/WavePrep**

***Index Terms*—intensive care, MIMIC, AI-ready, open source software, signal processing, biosignal**

## I. Introduction

The Medical Information Mart for Intensive Care (MIMIC) is a well-established and valuable resource for researchers working with data from intensive care units (ICUs). There are multiple versions of MIMIC, each containing data recorded over different time frames; the most recent is MIMIC-IV, published in 2023. The MIMIC databases II-IV consist of two parts: "MIMIC core" and "MIMIC waveform". The MIMIC core consists of a relational database containing various chart events, including vital signs and medications, as well as patient information. MIMIC waveform contains high-resolution bedside patient monitoring. Different levels of resolution are available: *numeric* waveforms contain recordings such as heart rate, oxygenation level, or blood pressure at 1-per-minute measurements; *non-numeric* waveforms contain higher-resolution recordings, such as plethysmography. The "MIMIC matched waveform" database allows retrieval of additional information about a patient's stay by matching $patient\_id$ to the recording date. These waveforms are stored in the Waveform Database (WFDB) format, which is widely used in the PhysioNet [1] datasets.

M. Elwes, O. Beyan and E. Kutafina are with the University of Cologne, Faculty of Medicine and University Hospital Cologne, Institute for Biomedical Informatics, Germany, e-mail: mayra.elwes@uk-koeln.de.

Despite widespread adoption of core MIMIC databases, MIMIC waveform databases are comparably underutilised. Although the core MIMIC-III [2] has 98.000 citations on Google Scholar as of December 2025, compared to 124 citations of the MIMIC-III waveform [3]. In contrast to core MIMIC, MIMIC waveforms are publicly available without the need to apply for credential access. One reason for the discrepancy between the use of the core MIMIC databases and the MIMIC (matched) waveform databases could be researchers' unfamiliarity with the WFDB format rather than with relational databases.

To use a dataset for machine learning, it first needs to be made machine learining-ready, or, more broadly *AI-ready*. This includes adequate preprocessing steps and proper splitting into training and test sets. Apart from adequate preprocessing, an AI-ready dataset should be accompanied by metadata that allows researchers to assess its suitability for a research question. These metadata should include basic information (data ownership, version, data type), data quality information, and data provenance, including a report on all applied processing steps [4]. Standard pre-processing steps for waveform data include down- or up-sampling, data cleaning, imputing missing values, and windowing. Depending on the use case, these steps can be combined in various ways and may differ across channels of the same recording. To our best knowledge, there is no published tool for creating a machine-learning-ready dataset from WFDB databases.

In this paper, we propose WavePrep, an open-source software that provides a flexible pipeline for creating AI-ready datasets from the *numeric* and *non-numeric* MIMIC-III waveform database. In particular, we provide a user-friendly JSON interface that allows users to define preprocessing steps with fine-grained, step-by-step, and channel-specific control, perform a train-test split, and enable automatic parallelisation to speed up dataset creation significantly.

## II. Methods

To use WavePrep, the user must provide two input files; an impression is given in Figure 1. Firstly, a CSV file specifying the selected records, along with the time span for each record that should be included in the resulting dataset. Secondly, a JSON file specifying the channels to extract from the record, the channel-specific pre-processing steps, the windowing, and the split into train, test, and validation sets.

WavePrep iterates over the channel-specific steps, as shown in Algorithm 1, and supports three categories of processing

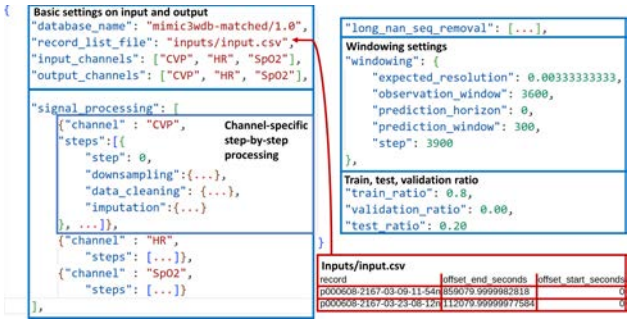


Fig. 1: Shortened exemplary config.json and input.csv, detailed descriptions are provided in the GitHub repository.

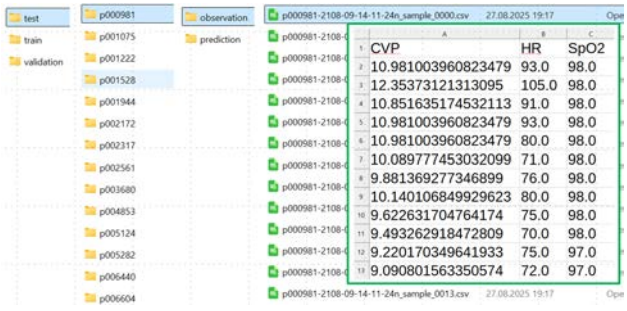


Fig. 2: Structure of WavePrep output.

steps: downsampling, data cleaning, and imputation. In particular, supported downsampling strategies include decimation, mean, and median downsampling. Data cleaning is supported in two ways. Firstly, filtering by replacing specific values with NaN. Currently, filtering is supported via channel-specific user-defined valid value ranges. Secondly, by removing sequences of the record in which at least one channel has a sequence of NaN values longer than a user-specified cut-off sequence length. A combination of filtering and removal of long NaN sequences can be used to remove sequences of implausible data, e.g., when a sensor was not connected to the patient. For imputation, only one strategy is currently supported: forward filling the first three missing values and then using an iterative imputer to fill the remaining values.

**Algorithm 1:** WavePrep - Signal Processing

```
Data: config, start_step, end_step, S
Result: S
for step_idx ← start_step to end_step do
    foreach channel ∈ specified_channels do
        for i ∈ len(S) do
            channel_data ← S[i][channel] if step_idx ∈
             config("channel_specific_steps") then
                channel_data ← chan-
                 nel_specific_steps[step_idx](channel_data,
                 config);
            end
        end
    end
    if step_idx ∈ config("remove_nan_sequences")
     then
        S ← remove_nan_sequences(S, config)
    end
end
```

WavePrep automatically determines when a preprocessing step, such as iterative imputation, requires defined train-test division, and will window and split the dataset before executing it. WavePreps algorithm is shown in Algorithm 2. Windowing is done based on four parameters: the duration of the observation window, the prediction horizon, the prediction window, and an offset to the next window. The dataset split is performed based on user-specified train-, validate- and test-ratios. To avoid data leakage, a group-shuffle split at the patient level is applied.

**Algorithm 2:** WavePrep - Overarching Process

```
Data: records_start_end, config
Result: processed samples in csv format
start_step, end_step, max_step ← get_steps(config)
 parallel for record, start_offset, end_offset ∈
 records_start_end do
    S ← fetch_wfdb_record_snippet;
    S ← signal_processing_(config, start_step,
     end_step, S);
    windows ← windowing(S, config);
    store_windows_to_disc(windows, config);
end
split(config);
if imputer needs to be trained then
    train imputer and store to disc;
end
if end_step ≤ max_step then
    for split_type ∈ { "train", "validate", "test"} do
        parallel for folder in folders in
         folders(split_type) do
            S ← fetch and transform data from disc;
            S ← signal_processing_(config, end_step,
             overall_max_steps, S);
            store_to_disc(S, config);
        end
    end
end
```

The output of WavePrep is CSV files of the observation and prediction windows, is organised in a specific file structure, as shown in Figure 2. At the first level, the resulting dataset is split into *train*, *validation*, and *test* sets. Within these folders, the processed samples are grouped by patient and named according to the rule *record_id_sample_number.csv*. All observation windows of a patient's recordings are grouped under that patient's *observation* folder; likewise, the prediction windows are grouped under the *prediction* folder. Within the CSV files, one column corresponds to one channel, and each row corresponds to a timepoint.

WavePrep is developed in Python 3.12.3. It mainly depends on WFDB, Pandas, NumPy, and scikit-learn. The preprocessing step categories are implemented as abstract classes, enabling researchers to extend the framework with specific implementations of these steps. Records are processed in

TABLE I: Performance of WavePrep.

| Data-set | # Records | Total duration of records [s] | Processing time [s] | Avg time / record [s] | Avg time / hour of record [ms] |
|---|---|---|---|---|---|
| 1 | 1092 | 78657.67 | 2603.75 | 2.38 | 33.10 |
| 2 | 945 | 54629.98 | 1853.87 | 1.96 | 33.94 |
| 3 | 456 | 42522.57 | 1595.46 | 3.50 | 37.52 |
| 4 | 512 | 36965.40 | 1430.72 | 2.79 | 38.70 |
| 5 | 565 | 35293.08 | 1414.06 | 2.50 | 40.07 |
| 6 | 445 | 21335.88 | 889.84 | 2.00 | 41.71 |
| 7 | 440 | 23041.47 | 968.62 | 2.20 | 42.04 |
| 8 | 241 | 14184.00 | 716.69 | 2.97 | 50.53 |

parallel.

## III. Results

We assess WavePrep's processing speed by creating eight datasets. All records are selected from the *numeric* records of MIMIC-III matched waveform v1.0. The selected records and time spans specified in input.csv differ across datasets, and the number of recordings and total duration to be processed for each dataset are shown in Table I. The preprocessing steps specified in config.json are the same for all datasets.

The preprocessing steps specify extracting data for six channels. Each of these channels is processed in the same way: (1) mean downsampling to 1/1 minute, (2) exclusion of data outside of a specified range, (3) imputation via an iterative imputer, and (4) finally mean downsampling to 1/5 minutes. In this configuration, WavePrep automatically detects that, to apply step (3), an imputer needs to be trained on the test dataset's observation windows. So it performs windowing and train-test split before step (3). For windowing, the recordings are sliced into 60-minute observation windows, 0-minute prediction horizons, and 5-minute prediction windows, in a non-overlapping manner.

The experiments are performed on a Server with two Intel Xeon Silver 4316 (40 cores, 80 threads @2.30GHz). The results of the experiments are shown in Table I. The average processing time per record is between 1.96 and 3.50 seconds, and per hour of recording, between 33.10 and 50.53 milliseconds. All encountered records were processed correctly.

Compatibility with the *non-numeric* records of MIMIC III matched waveforms was confirmed via preprocessing that included downsampling and data cleaning of the plethysmography channel over the first 60 seconds of eight records.

WavePrep was successfully used to prepare various datasets for a study benchmarking AI models for vital parameter prediction tasks [6], demonstrating its ability to create machine-learning-ready datasets.

## IV. Discussion

WavePrep generated the datasets for the MIMIC-III matched waveform v1.0 experiments as expected, demonstrating its ability to preprocess data to meet the needs of a machine learning project. By transforming data from WFDB to CSV, a format already familiar to many researchers, WavePrep may lower the barrier to working with the MIMIC waveform databases. The processing time per recorded hour decreases with increasing number of records to process, due to parallelisation in WavePrep. The open-source nature of the tool and the use of abstraction enable the community to adapt and further develop WavePrep to their specific preprocessing needs. The processing step categories are limited to downsampling, data cleaning, and imputing, for now, but can be expanded following the abstraction pattern to include further pre-processing categories. Using WavePrep addresses AI-readiness requirement through metadata by providing the config.json and input.csv files, which detail the processing and data split applied to the dataset.

## V. Conclusion

WavePrep is a tool for creating machine-learning-ready signal datasets from MIMIC waveform databases. The tool allows for configuring pre-processing in a flexible, fine-grained, and channel-specific manner, supporting common signal preprocessing steps, such as downsampling, data cleaning, cutting out long NaN-sequences, imputation, windowing, and splitting a dataset into training and test data. These steps can be specified via a user-friendly JSON file, lowering the entry barrier to using MIMIC waveform data. An exemplary performance analysis on a server with 80 available workers showed that the tool takes between 33.10 and 50.53 milliseconds to process 1 hour of recording when processing 6 channels. The modular, open-source tool enables researchers to add functionality according to their needs. In future work, we will extend WavePrep with a detailed dataset quality report, further addressing AI-readiness and helping researchers create high-quality signal datasets with minimal coding. Additionally, we will integrate the Croissant metadata format for datatsets [5], addressing AI-readiness demands of the machine learning community.

## References

[1] A. Goldberger, A. Amaral et.al. *PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals.* Circulation [Online], 2000.

[2] A. E. W. Johnson et al., *MIMIC-III, a freely accessible critical care database.* Sci Data, vol. 3, no. 1, p. 160035, May 2016. doi:10.1038/sdata.2016.35

[3] B. Moody, G. Moody, M. Villarroel, G. Clifford, and I. Silva, *MIMIC-III Waveform Database Matched Subset (version 1.0).*, PhysioNet. RRID:SCR_007345, 2020. doi: 10.13026/c2294b

[4] S. Afzal, C. Rajmohan, M. Kesarwani, S. Mehta, and H. Patel, *Data Readiness Report*. In: 2021 IEEE International Conference on Smart Data Services (SMDS), Chicago, IL, USA: IEEE, Sep. 2021, pp. 42–51. doi: 10.1109/SMDS53860.2021.00016

[5] M. Akhtar et al., *Croissant: A Metadata Format for ML-Ready Datasets.* In: Proceedings of the Eighth Workshop on Data Management for End-to-End Machine Learning. Santiago AA Chile: ACM; 2024. p. 1–6. doi: 10.1145/3650203.3663326

[6] M. Elwes et al., *Distribution Shift Analysis in Generalizable Modelling: Intensive Care Time-Series Data*, Submitted for publication at MIE 2026. doi: 10.13140/RG.2.2.14589.83684

 DOI 10.15203/99106-194-6-25

# Analysis of factors affecting the signal quality in imaging photoplethysmography

Fabienne Sahl, Vincent Fleischhauer, Alexander Woyczyk, Sebastian Zaunseder

***Abstract*—This study investigates the effects of sympathetic activation, induced by a cold pressor test (CPT), and varying body position (supine and upright) on the signal quality in imaging photoplethysmography (iPPG) gained from the face under controlled illumination conditions. We used own experimental data and assessed the signal quality by two common signal quality indices. Statistical analysis was done using linear mixed-effects models. Our results demonstrate a significant reduced signal quality in upright versus supine position. Sympathetic activation by CPT did not reduce the signal quality in iPPG (and in some cases, even increased it). As expected, the signal quality is reduced in iPPG compared to finger PPG but interestingly the behavior upon CPT differs, hinting at different dominant mechanisms in facial iPPG and finger PPG.**

***Index Terms*—imaging photoplethysmography, signal quality**

## I. Introduction

Photoplethysmography (PPG) is a well-established non-invasive measurement technique widely used in clinics and everyday applications. PPG signals, which have traditionally been acquired by using contact sensors at peripheral sites such as the fingertip, are susceptible to motion artifacts and external disturbances. The contactless variant, imaging photoplethysmography (iPPG), is even more sensitive to motion-related artifacts, necessitating robust motion compensation and tracking algorithms. Such disturbances can substantially degrade signal quality and compromise the reliability of derived physiological parameters as multiple studies have shown [4]. However, few studies have done systematic research on factors affecting the signal quality in iPPG.

In this study, we examine the effects of sympathetic activation, induced by a cold pressor test (CPT), and varying body position (supine and upright) on the signal quality in iPPG.

## II. Material & Methods

### A. Data

The analyses presented in this work are based on own experimental data. The dataset used comprises synchronized recordings of contact-based PPG (finger and earlobe), iPPG, and other biosignals, collected from 43 subjects (29 male, 14 female; age range 20–59 years) undergoing different stimuli. Each recording has a duration of approximately 50 minutes.

iPPG signals were derived from facial video recordings acquired using an RGB camera (UI-3060CP-C-HQ Rev. 2, IDS Imaging Development Systems GmbH, Germany). The camera was mounted with a fixed orientation toward the subject at a distance of approximately 40 cm and recorded the head and upper shoulder region. Videos were captured at a spatial resolution of 1280 × 960 pixels, a frame rate of 25 fps, and a color depth of 12 bit under constant artificial illumination.

F. Sahl, A. Woyczyk and S. Zaunseder are with the University of Augsburg, Chair for Diagnostic Sensing, Augsburg, Germany, fabienne.sahl@uni-a.de.

V. Fleischhauer is with the University of Augsburg, Chair of Neurorehabilitation, Germany.

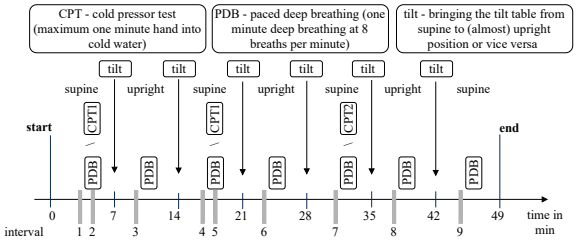


Fig. 1: Experimental protocol. CPT1 was randomly assigned to the first or second supine position. Subjects undergoing CPT2 were randomly assigned. Whenever no CPT is executed, PDB is done. Intervals 1–9 denote the segments from which PPG signals were extracted for this study.

Contact-based PPG signals were recorded at 2000 Hz sampling frequency using reflective photoplethysmographic sensors with a wavelength of 860 ± 60 nm attached to the fingertip and the earlobe, connected to a biosignal acquisition system (Biopac MP36, Biopac Systems Inc., USA). All signals were resampled to a sampling frequency of 100 Hz.

The experimental protocol included interventions known to affect perfusion: a cold pressor test (CPT), paced deep breathing (PDB), and multiple postural changes between supine and upright positions. Six tilt transitions were performed throughout the experiment, starting in supine position. A CPT (CPT1) was applied randomly in the first or second supine interval. Some randomly selected subset of subjects also underwent a second CPT (CPT2) in the third supine interval. PDB was performed during phases without a CPT. Figure 1 shows the complete experimental protocol.

### B. Preprocessing

Five methods to derive iPPG signals were considered, along with two contact PPG signals acquired at the finger and earlobe. An overview is provided in Table I. The contact PPG signals were used as recorded.

For iPPG extraction, facial regions were segmented using Gaussian mixture models and level-set methods, following the approach described by Woyczyk et al. [6].

To compare the signal quality across different stimuli, nine 20 s signal segments (interval 1-9) were extracted as shown in figure 1. The signals during CPT were extracted starting 30 s before the end of CPT1 (interval 2 or 3). All other

TABLE I: Description of extracted (i)PPG signals.

| short name | type | Description |
|---|---|---|
| Green | imaging | green channel |
| nGreen | imaging | DC normalized green channel |
| OPTACC | imaging | combination of RGB channels [1] |
| POS | imaging | combination of RGB channels (plane orthogonal to skin) [5] |
| CHROM | imaging | combination of RGB channels (chrominance based) [3] |
| fPPG | contact | contact PPG measured at finger |
| ePPG | contact | contact PPG measured at earlobe |

signals were extracted starting 30 s before CPT or PDB. Before further processing, all PPG segments were mean-centered. The demeaned signal is denoted as $\tilde{x}$ and was subsequently band-pass filtered between 0.5 Hz and 5 Hz yielding $x^{bp}$ as signals of interest.

To assess the effect of sympathetic activation we compare interval 1 versus 2 or 4 versus 5, depending on where the CPT1 was executed. Only CPT1 was considered, as it was performed by all subjects and always occurred in the first half of the experimental protocol.

To assess the effect of body position, intervals 1, 3, 4, 6, 7, 8 and 9 are used, yielding a total of four PPG segments in supine (1, 4, 7, and 9) and three segments in upright position (3, 6 and 8). These time points were assumed to be least affected by other interventions and thus best reflect positional effects.

To analyse the longitudinal change of signal quality, the same intervals as for the analysis of body position were used.

### C. Signal quality indices

Various SQIs exist for PPG assessment [4]. In this study, we expect physiological effects (e.g. changes in PPG amplitude) between the compared conditions. Therefore, we selected two SQIs that we assume to be minimally influenced by the expected physiological effects.

*a) Signal-to-noise ratio (SNR):* The SNR is calculated using frequency filtering. The demeaned signal $\tilde{x}$ was high-pass filtered with a cutoff frequency at 0.5 Hz resulting in $x^{hp}$, which contains the physiological PPG component and high-frequency noise. $\tilde{x}$ was then low-pass filtered with a cutoff frequency of 5 Hz, resulting in $x^{lp}$, which contains the physiological component and low-frequency noise. SNR is then derived using equation 1 with $n$ being the number of samples in a PPG segment.

$$SNR = 10\log\left(\frac{\sum_{k=1}^{n}\left(x_k^{bp}\right)^2}{\sum_{k=1}^{n}\left(x_k^{bp}-x_k^{hp}\right)^2+\sum_{k=1}^{n}\left(x_k^{bp}-x_k^{lp}\right)^2}\right) \quad (1)$$

*b) Beat morphology correlation (corr):* corr is a measure that compares the morphology of single beats within each segment. After peak detection in the filtered PPG signal $x_k^{bp}$, cardiac cycles were segmented, and corr was obtained by averaging pairwise Pearson correlation coefficients.

### D. Statistics

Linear mixed-effects models (LMM) with subject-specific random intercepts were used to analyse repeated measurements across different conditions. The analysis aimed to investigate (1) changes in signal quality induced by CPT and (2) changes in signal quality between lying and standing positions. To consider additional effects which might be introduced by our experimental design, we additionally analysed temporal changes in signal quality across measurement intervals throughout the whole recording duration, both overall and separately for lying and standing conditions as a third factor. Notably, such analysis specifically directs to our experiments, while the other analyses have a generalized claim.

SNR and corr were analysed separately. To enable statistical modeling for corr, the values were transformed to an unbounded scale using Fisher's $z$-transformation. Temporal effects were modeled using numerically encoded measurement intervals. We tested against a p-value $< 0.05$. For both research questions, effect sizes were estimated using Hedges' $g$. $g > 0$ indicates a decrease in SQI while $g < 0$ indicates an increase in SQI.

## III. Results

TABLE II: Results from LMMs comparing SQIs for pre-CPT vs. during-CPT conditions and for supine vs. upright positions. Significant differences are highlighted in light gray.

| | **CPT** | | | | **posture change** | | | |
|---|---|---|---|---|---|---|---|---|
| | **SNR** | | **corr** | | **SNR** | | **corr** | |
| **PPG** | $p$ | $g$ | $p$ | $g$ | $p$ | $g$ | $p$ | $g$ |
| CHROM | 0.09 | −0.17 | 0.07 | −0.15 | 0.00 | 0.50 | 0.00 | 0.44 |
| ePPG | 0.23 | 0.11 | 0.93 | −0.01 | 0.00 | 0.41 | 0.02 | 0.17 |
| fPPG | 0.03 | 0.38 | 0.05 | 0.25 | 0.01 | 0.28 | 0.80 | 0.02 |
| Green | 0.01 | −0.31 | 0.51 | −0.06 | 0.01 | 0.27 | 0.00 | 0.47 |
| nGreen | 0.00 | −0.36 | 0.50 | −0.06 | 0.00 | 0.30 | 0.00 | 0.48 |
| OPTACC | 0.30 | −0.12 | 0.11 | −0.13 | 0.00 | 0.40 | 0.00 | 0.35 |
| POS | 0.07 | −0.18 | 0.11 | −0.13 | 0.00 | 0.43 | 0.00 | 0.29 |

Table II summarizes p-values and effect sizes of the used SQI for CPT and posture-related effects. During CPT, SNR showed significant changes for fPPG, Green, and nGreen, whereas corr changed significantly only for fPPG. No other significant effects were observed for the CPT condition.

Postural changes had a stronger impact on signal quality, with significant effects observed for almost all PPG signals and SQIs. Figure 2 presents boxplots on the SQI behaviour for the CPT condition across all PPG signals.

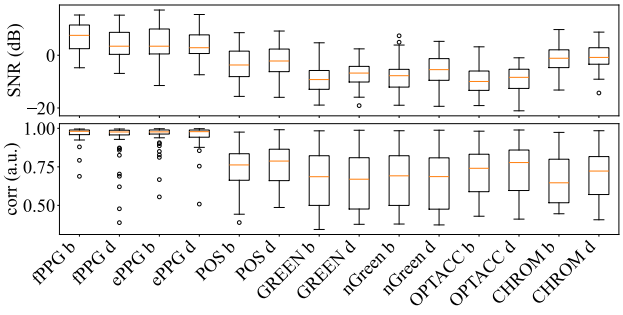


Fig. 2: Distribution of corr and SNR across all subjects for each PPG type before (b) and during (d) CPT.

Table III shows the p-values of the longitudinal analysis. When considering time effects irrespective of body position,

 DOI 10.15203/99106-194-6-25

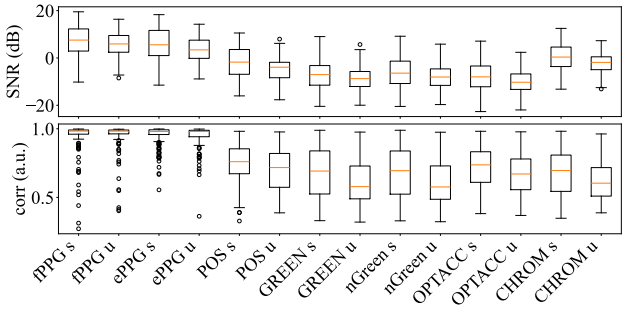


Fig. 3: Distribution of corr and SNR across all subjects for each PPG type in supine (s) and upright (u) position.

TABLE III: Results from LMMs comparing SQIs for longitudinal effects without considering postures, longitudinal effects in upright position and longitudinal effects in supine position. Significant differences are highlighted in light gray.

| | longitudinal | | longitudinal upright | | longitudinal supine | |
|---|---|---|---|---|---|---|
| **PPG type** | **SNR** | **corr** | **SNR** | **corr** | **SNR** | **corr** |
| CHROM | 0.00 | 0.00 | 0.16 | 0.00 | 0.00 | 0.00 |
| ePPG | 0.00 | 0.00 | 0.07 | 0.04 | 0.00 | 0.00 |
| fPPG | 0.00 | 0.10 | 0.34 | 0.35 | 0.00 | 0.13 |
| Green | 0.00 | 0.00 | 0.32 | 0.00 | 0.00 | 0.00 |
| nGreen | 0.00 | 0.00 | 0.38 | 0.01 | 0.00 | 0.00 |
| OPTACC | 0.00 | 0.00 | 0.04 | 0.00 | 0.00 | 0.00 |
| POS | 0.00 | 0.00 | 0.10 | 0.00 | 0.00 | 0.00 |

significant changes were observed over time for nearly all combinations of SQI and PPG modality.

The longitudinal analysis by posture revealed that corr behaved consistently across postures: all iPPG modalities exhibited significant longitudinal changes, whereas contact fPPG signals did not. In contrast, SNR showed a posture-dependent longitudinal behavior. In upright position, except for OPTACC, there were no significant longitudinal effects. Conversely, in supine position, SNR exhibited significant longitudinal changes for all remaining PPG modalities. These effects are illustrated in Figure 4, where the final interval shows lower values compared to earlier intervals. Additionally, posture-related differences reported in Table II can be visually inferred from the figure.

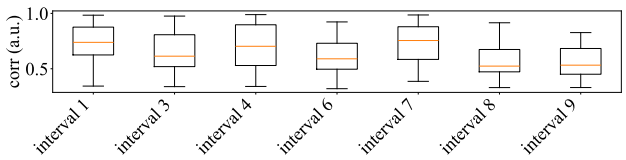


Fig. 4: Distribution of corr across all subjects for POS per interval.

## IV. Discussion

As expected, iPPG signal quality is generally lower compared to contact PPG signal quality. Its good recording conditions make fPPG morphology robust against tilting. Concerning the iPPG variants, POS and CHROM offer the highest signal qualities. Regarding the considered factors, our results indicate a pronounced effect on iPPG quality owing to body position where the signal quality systematically decreases in upright position. Such a reduced quality might be due to a more unstable position of the head in upright position. Sympathetic activation does typically not lead to significant changes in signal quality in our experiments. While vasoconstriction might be linked to reduced pulsation and reduced signal quality, interestingly, the signal quality tends to be even higher for iPPG upon CPT, probably due to an increased pulse pressure as a result of systemic regulation. The significant drop observed in SNR and corr for fPPG may be caused by pronounced vasoconstriction during CPT. The finger is expected to be more affected as it is a more peripheral site than the face, as our earlier analyses on the signals' shape also indicated [2].

The significant effects observed in the longitudinal and longitudinal–supine analyses were primarily driven by the final interval 9. This effect was consistently observed across multiple PPG modalities. Its origin is difficult to explain. As finger PPG and iPPG are affected, the finding does not seem to be related to the automated iPPG processing. It might be related to more movements owing to increasing discomfort after an extended experimental time.

## V. Conclusion

We analysed SQIs on multiple iPPG modalities upon different stimuli. CPT had little effect on signal quality, although SNR may decrease slightly due to physiological changes. Body posture significantly influenced signal quality, with better signals observed in the supine position. Stimulus-related effects in contact PPG were mainly reflected in SNR, while signal morphology remained stable, likely due to the direct measurement on the skin and the inherent robustness of the technique. Longitudinal analysis revealed a gradual decrease in SQI over time, suggesting that further investigation is needed to rule out systematic errors in PPG processing. Moreover, the red and blue color channels of iPPG as well as the paced deep breathing sequences will be considered in future SQI analyses.

## References

[1] Hannes Ernst et al. "Optimal color channel combination across skin tones for remote heart rate measurement in camera-based photoplethysmography". In: *Biomedical Signal Processing and Control* 68 (Apr. 22, 2021).

[2] Vincent Fleischhauer et al. "Photoplethysmography upon cold stress—impact of measurement site and acquisition mode". In: *Frontiers in Physiology* 14 (June 1, 2023).

[3] Gerard de Haan and Vincent Jeanne. "Robust pulserate from chrominance-based rPPG". In: *IEEE transactions on bio-medical engineering* 60.10 (2013).

[4] Junyung Park et al. "Photoplethysmogram Analysis and Applications: An Integrative Review". In: *Frontiers in Physiology* 12 (Mar. 1, 2022).

[5] Wenjin Wang et al. "Algorithmic Principles of Remote-PPG". In: IEEE, 2016.

[6] Alexander Woyczyk, Vincent Fleischhauer, and Sebastian Zaunseder. "Adaptive Gaussian Mixture Model Driven Level Set Segmentation for Remote Pulse Rate Detection". In: *IEEE Journal of Biomedical and Health Informatics* 25.5 (May 2021).

 DOI 10.15203/99106-194-6-26

# Three sensor OPM-MEG setup to detect the auditory M100 response

A. Jodko-Władzińska, P. Tappayuthpijarn, T. Sander

***Abstract*—Magnetoencephalography with optically pumped magnetometers is rapidly evolving both for full head coverage applications and for low sensor count applications. Here a three dual-axis OPM setup is introduced to measure auditory brain responses. Without specific anatomical knowledge about the participant an auditory M100 response was detected with three sensors over the left temporal lobe for a sound delivered into the contralateral right ear. An accelerometer was used to investigate the effect of unavoidable participant movements on the resulting average signals.**

***Index Terms*—magnetoencephalography, optically pumped magnetometer, auditory response**

## I. Introduction

Magnetoencephalography based on optically pumped magnetometers (OPM-MEG) allows completely new setups tailored to specific needs and potentially simplifying investigations. Auditory steady state responses were observed with only six OPMs in [1] and emotionally charged visual stimuli were investigated by nine OPMs in [2]. With fifteen OPM sensors source localization was possible for auditory responses in a group study [3] and potentially diagnostically relevant information from magnetic heart signals was obtained with fourteen OPMs in [4]. This is possible since most OPMs nowadays detect two components of the magnetic vector field and for fifteen sensors already 30 independent signals are obtained.

Here we investigate a further simplification by using three dual-axis sensors made by Q-MAG Technology (Hangzhou, China) and a magnetic coil to generate trigger pulses directly into the OPM signals. For this feasibility study one participant is measured while listening to tones at various volumes below and above the individual hearing threshold.

## II. Methods

To record MEG from the auditory cortex three dual-axis OPMs manufactured by Q-mag Technology Hangzhou were inserted into a 3D-printed helmet fitting the head of the participant. Helmet and OPMs are shown in Fig. 1 and the position of the OPMs is over the left temporal region of the participant and about 3 cm above the left ear canal. The ipsilateral side was not measured due to the limited number of sensors available. A sequential measurement of ipsi- and contralateral side is possible, but conditions might change between the two sessions.

Tones with a frequency of 1 kHz and a duration of 400 ms were presented to the right ear using an ER-30 (etymotic.com, discontinued product) non-magnetic earphone at five different sound pressure levels: -3 dB below hearing threshold (HT), at HT, +6, +12, and +30 dB above HT (= 30 dB sensational level = SL). The levels were played in random order with 240 repetitions each. The HT was determined in the same experimental setting just before the MEG recording. At the start of each tone a 30 ms-long trigger pulse was generated and fed to the coil shown in Fig. 1 (left inset), which serves the purpose of generating a magnetic trigger signal.

The sensors stream their data over a network connection and a network switch in real time to a PC, but there is no analog trigger signal input in the sensor electronics to record a voltage signal. Therefore, a magnetic trigger solution was implemented in this proof-of-concept study. The coil is about 1 m away from the helmet and feeding a current through the coil generates a signal in the sensors. Since this coil was energized at the same time as the tone started the magnetic sensor signal contains the trigger pulses as can be seen in Fig. 2.

The second inset in Fig. 1 (top right) shows a triple axis accelerometer (www.pluxbiosignals.com) attached to the apex of the helmet. The accelerometer signal is recorded over a Bluetooth wireless connection provided by the manufacturer and was tested to not interfere with OPM operation. This electronics contains an analog input, and the trigger signal was recorded synchronously with the accelerometer signal.

During the experiment the participant was seated in a chair and the head position was stabilized by a neck rest.

Fig. 1. Participant helmet with three dual-axis OPMs inserted over the auditory cortex. Insets show (top left) the coil used for generating the trigger pulses and (top right) the triple-axis accelerometer attached at the apex of the helmet.

A. Jodko-Władzińska is with Warsaw University of Technology, Warsaw, Poland.

P. Tappayuthpijarn and T. Sander are with Physikalisch-Technische Bundesanstalt, Berlin, Germany. E-Mail: tilmann.sander-thoemmes@ptb.de.

 DOI 10.15203/99106-194-6-26

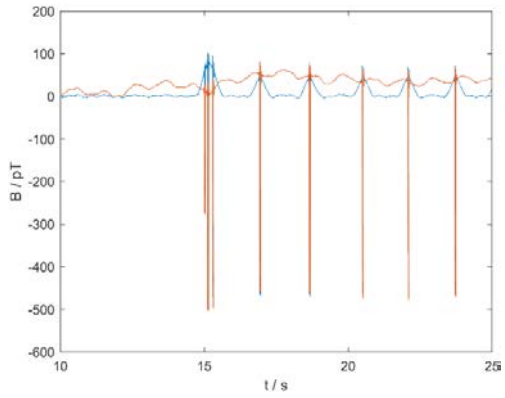


Fig. 2. Trigger signal generated by the coil inside the magnetically shielded room shown in the left inset in Fig. 1. The orange curve is the raw signal from an OPM and the blue curve is high pass filtered at 1 Hz to remove the background. The trigger peaks are clearly visible.

After the recording the OPM magnetic signals and accelerometer data were averaged using the trigger time points detected in the OPM signals shown in Fig. 2. Before averaging the magnetic signals were bandpass filtered between 1 and 45 Hz. The sequence of the different stimuli was random and recorded in a log-file of the Presentation software (www.neurobs.com).

## III. Results

The average radial magnetic signal of the three OPMs is shown in Fig. 3a) for the +6 dB SL and in b) for the +30 dB SL tone volume. The oscillatory signal starting at t = -50 ms is the average trigger pulse. As theoretically expected, the oscillations ("ringing") show that the 75 Hz bandwidth of the OPM cannot resolve the rectangular trigger pulse with an amplitude of several 100 pT observable in Fig. 2. The strong trigger does not saturate the OPM, and the oscillations cease around 75 ms and the window from 75 to 85 ms is chosen for the baseline correction.

Usually, the baseline window is chosen before the tone onset, but the strong sensor bandwidth and low pass filter induced "ringing" of the pulse in Fig. 3 necessitated a baseline window just before the M100 onset. Furthermore, there is a level shift before and after the pulse in Fig. 3a) probably due to a slow signal background not removed by the 1 Hz high pass filter. Clearly brain responses before the M100 are suppressed by this window choice necessary for these pilot data. The occurrence of oscillations before the rising edge of the pulse in Fig. 3 suggests a symmetric filter in the sensor software.

For the +30 dB SL tone volume in Fig. 3b) a clear downward peak is observed at 120 ms in two adjacent sensors, but not in the third sensor. This indicates that the peak is from a source close such as the auditory cortex and the amplitude of ~700 fT is expected for an auditory M100 brain response. The exact time of the M100 occurrence can vary between participants. Furthermore, a sensor processing delay of 10 ms is likely due to the bandwidth of 75 Hz as stated by the manufacturer. In summation the 120 ms occurrence is probably an individual physiological response time of 110 ms combined with a 10 ms internal sensor delay.

After 200 ms the signal is similar for all three OPMs indicating that it is due to a distant source. For the +6 dB SL tone volume no peak is observed, but again after 200 ms the signal is similar for all three OPMs.

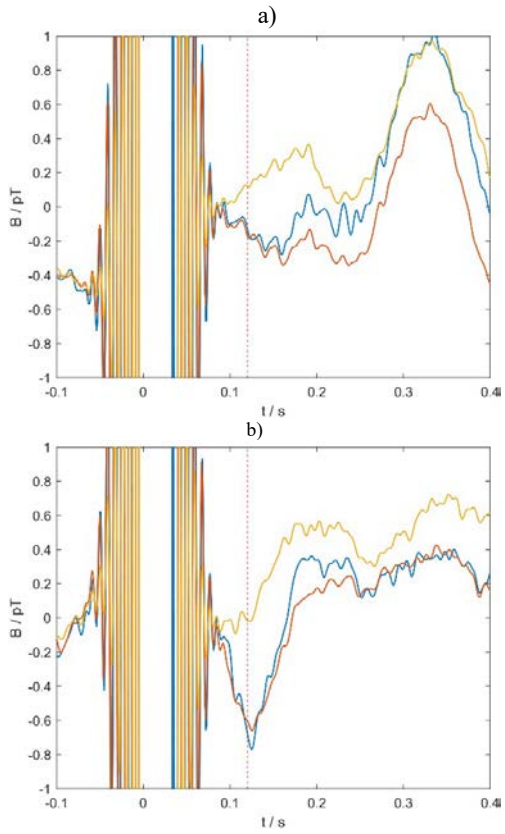


Fig. 3. Averaged magnetic signal for the +6 (a) and +30 dB (b) SL tone volume for the radial direction signal of the three OPMs. At +30 dB tone volume a (negative) peak is observed at 120 ms in two sensors.

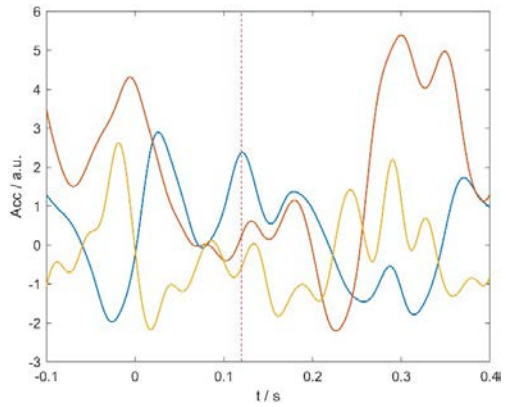


Fig. 4. Averaged xyz-accelerometer signals for the trigger points of the +30 dB stimuli. No obvious similarity exists to the magnetic signals in Fig. 3b).

To better understand the OPM signals the accelerometer data were averaged as well. Participant movements can induce common signals in all OPMs due to the gradients in a magnetically shielded room which are spatially homogeneous in the center of the room for the typical volume of 30x30x30 cm^3 occupied by the OPM-MEG array. The average accelerometer signal for the +30 dB condition is shown in Fig. 4 with separate curves for the xyz-outputs. The shape of the accelerometer signals in Fig. 4 does not show the almost flat OPM signal average for t > 200 ms in Fig. 3b). The unavoidable participant movements are apparently not causing strong additional noise.

## IV. Conclusion

An auditory M100 response was clearly observable in the three OPM / six channel setup for a sound pressure level of +30 dB SL, although it was not visible at a volume of + 6 dB SL. Future work will increase the number of stimulations to improve SNR for applications in hearing research. The triggering using a magnetic signal recorded by the OPMs simplifies the hardware, but the amplitude of the pulse needs to be reduced to allow an easier identification of brain signals. These OPM sensors have separate analog outputs and future work will use these outputs together with a multichannel DAQ to record sensor signals and trigger signal together. This will avoid the "ringing" in the magnetically recorded trigger signal. Future work will replace the MRI based sensor helmet design by an approach using a cost effective 3D scanning method.

## Acknowledgment

The authors would like to thank Qiang Lin, Zhejiang University of Technology, and Huang Yuxiang, Hangzhou Q-MAG Technology Co.

## References

[1] An, Km., Shim, J.H., Kwon, H. *et al.* Detection of the 40 Hz auditory steady-state response with optically pumped magnetometers. Sci Rep 12, 17993 (2022).

[2] Jodko-Władzińska A., Sander T. Emotionally Charged Visually Evoked Magnetic Fields., Acta Phys Pol A 146(4), 521 (2024):

[3] Marhl, U., Jodko-Władzinska, A., Brühl, R., Sander, T., Jazbinšek, V.: Transforming and Comparing Data between Standard SQUID and OPM-MEG Systems. PLoS ONE 17, e0262669 (2022).

[4] Su S, Xu Z, He X, Zhang G, Wu H, Gao Y, Ma Y, Yin C, Ruan Y, Li K, Lin Q. Vector magnetocardiography using compact optically-pumped magnetometers. Heliyon 10(7), e29092 (2024).

 DOI 10.15203/99106-194-6-27

# TRANSFER
# Prediction of intraoperative Hypotension

O. Kahriman, A. Sander, S. Treskatsch, N. Wessel

***Abstract*—Intraoperative hypotension (IOH) is associated with serious postoperative complications including myocardial infarction, acute kidney injury, and increased mortality. Whilst commercial prediction algorithms exist, recent studies have questioned their validity, suggesting they rely predominantly on simple mean arterial pressure thresholds. We present a machine learning framework for predicting both the severity and endotype of impending hypotensive events using non-invasive measurements. Our approach classifies IOH into six severity levels and predicts the physiological endotype (myocardial depression, bradycardia, vasodilation, hypovolaemia, or mixed) to enable targeted therapeutic interventions. The models process multi-channel time series data (ECG, blood pressure, photoplethysmography, capnography) sampled at 512 Hz over a 5 min window to predict conditions 3 min ahead. Convolutional neural networks encode temporal patterns, with fully connected layers performing classification. Preliminary results on a subset of 180 cases from the VitalDB dataset show moderate discriminative ability (severity AUROC 0.64, endotype AUROC 0.58) but highlight challenges including severe class imbalance and limitations in endotype labelling methodology. Full-dataset training and baseline comparisons are in progress.**

***Index Terms*—blood pressure, convolutional neural networks, endotypes, hypotension prediction, intraoperative hypotension, machine learning, severity classification**

## I. Introduction

The maintenance of patients' vital signs within physiological ranges represents a fundamental responsibility of anaesthesiologists during surgical procedures. Patients who experience prolonged periods outside these ranges develop significantly more postoperative complications. For example, extended episodes of intraoperative hypotension (IOH) are associated with myocardial infarction, acute kidney injury, delirium, and death [1], [2], [3], [4], [5]. Postoperative complications result in prolonged hospital stays and substantial economic burden [6]. Thus, preventing intraoperative hypotension and its associated complications is critical for improving postoperative outcomes and patient wellbeing.

The current standard of care in the operating room involves continuous monitoring of patient vital signs with threshold-based alarms that alert clinicians when predefined limits are exceeded. However, physicians must continuously analyse these vital signs to anticipate future changes in patient status, adding substantial cognitive load during critical phases of anaesthesia. Our research project TRANSFER aims to address this challenge by integrating data from multiple sources with machine learning models to predict hypotensive events in advance, providing anaesthesiologists with critical lead time to intervene.

O.K. and A.S. are with ID Berlin, Germany, e-mail: o.kahriman@id-berlin.de.
N.W. is with MSB Medical School Berlin, Germany
S.T. is with the Charité – Universitätsmedizin Berlin, Germany

### A. Hypotension Prediction Index

The Hypotension Prediction Index (HPI) by Edwards Life Sciences predicts hypotension events 5 to 15 minutes in advance [7]. However, recent studies have raised concerns about its validity. The original validation study appears to suffer from selection bias [8], and direct comparisons showed similar AUROC between HPI and simple mean arterial pressure threshold-based predictors, suggesting HPI may rely predominantly on MAP rather than arterial waveform features [9]. These limitations highlight the need for alternative approaches in perioperative hypotension prediction.

### B. Severity Classes

IOH is commonly defined as MAP below 65 mmHg. However, predicting hypotension using this single threshold appears trivial, as simple MAP-based predictors achieve similar performance to complex models [9], [10]. Furthermore, IOH defined solely by the 65 mmHg threshold is not consistently associated with worse patient outcomes [11]. We propose a severity-based classification approach with six severity classes based on MAP thresholds: Class 0 ($\geq$75 mmHg), Class I ($<$70 mmHg), Class II ($<$65 mmHg), Class III ($<$60 mmHg), Class IV ($<$55 mmHg), and Class V ($<$50 mmHg). Training models to distinguish between these severity classes may provide more clinically relevant predictions than binary hypotension classification.

### C. Endotypes

Intraoperative hypotension events can arise from multiple physiological mechanisms requiring different therapeutic interventions. A 2023 study identified six distinct IOH endotypes through cluster analysis: myocardial depression, bradycardia, vasodilation with increased cardiac index, vasodilation without increased cardiac index, hypovolaemia, and mixed type [12]. These clusters were differentiated based on stroke volume index, heart rate, cardiac index, systemic vascular resistance index, and pulse pressure variation. We use the study's characteristic mean values and standard deviations for each endotype as reference distributions for labelling our training examples, aiming to predict not only the occurrence but also the endotype of impending IOH events for more targeted therapeutic interventions.

### D. Invasive and Non-invasive Measurements

Research and analysis of blood pressure typically relies on invasively measured arterial blood pressure. However, the risks associated with invasive arterial catheterisation are not justified in all surgical cases. We focus on predicting outcomes based on non-invasive measurements to enable broader applicability across all surgical procedures.

## II. Methods

For our analysis, we use the public VitalDB dataset, which contains over 6,000 surgical cases with continuous physiological signals and patient data [13]. Cases are split into training, validation, and test sets at the patient level to prevent data leakage. For this preliminary analysis, we utilised a subset of 180 randomly selected cases (108 training, 36 validation, 36 test).

Both tasks predict conditions 3 min into the future based on a 5 min observation window. Training examples are restricted to those with at most 10 s outside severity class 0 during the observation period. We define six severity classes based on MAP thresholds: Class 0 ($\geq$75 mmHg), Class I ($<$70 mmHg), Class II ($<$65 mmHg), Class III ($<$60 mmHg), Class IV ($<$55 mmHg), and Class V ($<$50 mmHg). For severity classification, the dataset contains equal numbers of hypotensive (classes I–V) and non-hypotensive (class 0) examples. Examples are labelled by time-averaging MAP between 3–4 min after $t_0$ and assigning the severity class with maximum duration in that interval.

For endotype prediction, the dataset contains only hypotensive examples (classes I–V). IOH events are labelled by calculating z-scores for each endotype using reference means and standard deviations from [12]. The endotype with maximum z-score is assigned as the label.

Models receive non-invasively measured time series data: ECG, systolic/diastolic/mean blood pressure, photoplethysmography, and capnography. All channels are resampled to 512 Hz for uniform temporal resolution. Target labels derive from invasively measured arterial blood pressure to ensure high-quality ground truth.

The models consist of four 1D convolutional layers with progressively increasing filters (32→64→128→256), decreasing kernel sizes (25, 15, 11, 7), stride-4 downsampling, batch normalization, ReLU activation, and 30 % dropout, followed by global average pooling, a fully connected layer and a softmax output for classification. The input is a segment of 5 min with 6 channels and a samplerate of 512 Hz. The models are implemented in Python using PyTorch. Severity classification is benchmarked against MAP-based baseline predictors, whilst endotype prediction is evaluated using one-vs-rest AUROC.

## III. Results

For this preliminary analysis, we utilised a subset of 180 cases from the VitalDB dataset (108 training, 36 validation, 36 test) to develop initial models, with full-dataset training planned prior to the conference.

TABLE I
Severity Classification Performance Metrics

| Class | Precision | Recall | F1 | AUROC | Support |
|---|---|---|---|---|---|
| 0 | 0.80 | 0.30 | 0.43 | 0.71 | 108 |
| I | 0.00 | 0.00 | 0.00 | 0.50 | 30 |
| II | 0.00 | 0.00 | 0.00 | 0.41 | 25 |
| III | 0.13 | 0.82 | 0.23 | 0.75 | 17 |
| IV | 0.00 | 0.00 | 0.00 | 0.71 | 5 |
| V | 0.58 | 0.29 | 0.39 | 0.77 | 24 |
| **Macro Avg** | 0.25 | 0.24 | 0.17 | 0.64 | 209 |
| **Weighted Avg** | 0.49 | 0.25 | 0.29 | 0.66 | 209 |

### A. Severity Classification

Table I shows the classification performance across all six severity classes and the AUROC for each class using a one-vs-rest approach.

The model achieved 25% accuracy on the multi-class classification task. Performance varied substantially across severity classes. Classes I, II, and IV achieved zero precision and recall. Class III showed the highest recall (0.82) but low precision (0.13), indicating over-prediction. Class 0 achieved moderate precision (0.80) but low recall (0.30), whilst Class V showed balanced but modest performance (precision 0.58, recall 0.29).

AUROC analysis revealed more encouraging discrimination between classes. The macro-averaged AUROC was 0.64, with Classes 0, III, IV, and V exceeding 0.70. Classes I and II showed AUROC values near chance (0.50 and 0.41, respectively).

### B. Endotype Classification

We trained a separate model on hypotensive segments (severity classes I–V) to classify physiological mechanisms. Table II presents classification performance across eight endotype categories and corresponding AUROC values.

The model achieved 12% accuracy on the validation set. Class 2 (no clear endotype match) accounted for 48% of validation samples (43 of 89) but achieved perfect precision (1.00) with very low recall (0.05). Class 1 (vasodilation with CI increase) showed high recall (0.69) but low precision (0.16). The remaining endotypes achieved zero precision and recall. Several classes had minimal representation: Class 7 (n=1), Class 6 (n=2), and Class 4 (n=4).

AUROC analysis revealed moderate discrimination despite poor classification (macro-average 0.58). Class 6 (hypovolaemia) achieved the highest AUROC (0.80), followed by Class 7 (bradycardia, 0.72) and Class 5 (0.68), though these estimates are unreliable given extremely small sample sizes. Classes 1 and 2 showed AUROC near chance (0.47 and 0.46).

## IV. Discussion

Our preliminary results demonstrate substantial challenges in predicting hypotension severity and endotypes from non-invasive measurements. The severity classification model achieved 25% accuracy with macro-averaged AUROC of 0.64, whilst endotype prediction achieved only 12% accuracy with AUROC of 0.58.

TABLE II
ENDOTYPE CLASSIFICATION PERFORMANCE METRICS

| Endotype | Prec. | Rec. | F1 | AUROC | Supp. |
|---|---|---|---|---|---|
| Class 1 (Vasodil. + CI) | 0.16 | 0.69 | 0.25 | 0.47 | 13 |
| Class 2 (No clear match) | 1.00 | 0.05 | 0.09 | 0.46 | 43 |
| Class 3 (Mixed type) | 0.00 | 0.00 | 0.00 | 0.60 | 17 |
| Class 4 (Myocard. depr.) | 0.00 | 0.00 | 0.00 | 0.31 | 4 |
| Class 5 (Vasodil. w/o CI) | 0.00 | 0.00 | 0.00 | 0.68 | 9 |
| Class 6 (Hypovolaemia) | 0.00 | 0.00 | 0.00 | 0.80 | 2 |
| Class 7 (Bradycardia) | 0.00 | 0.00 | 0.00 | 0.72 | 1 |
| **Macro Average** | 0.14 | 0.09 | 0.04 | 0.58 | 89 |
| **Weighted Average** | 0.51 | 0.12 | 0.08 | 0.52 | 89 |

For severity classification, the model showed systematic biases: over-predicting Class III (recall 0.82, precision 0.13) whilst under-predicting Class 0 (recall 0.30, precision 0.80). Classes I, II, and IV showed complete prediction failure. The discrepancy between moderate AUROC values (exceeding 0.70 for four classes) and poor classification accuracy suggests the model learns discriminative features but struggles with multi-class decision boundaries. Severe class imbalance (5–108 examples per class) and the difficulty of distinguishing 5 mmHg threshold differences from non-invasive measurements likely contribute to poor performance.

Endotype prediction proved more challenging. The prevalence of Class 2 (no clear match, 48% of segments) suggests systematic issues with our z-score-based labelling methodology. Extreme class imbalance (some classes with $n<5$) renders performance estimates unreliable. These results question whether non-invasive signals contain sufficient information to infer physiological mechanisms, as the relationship may be too patient-specific for generalizable predictions.

Several limitations affect interpretation. Our subset of 180 cases may be insufficient for deep learning. We have not yet implemented the MAP-based baseline comparison, which is essential for determining whether complex models provide value beyond simple threshold-based approaches that have shown strong performance in prior work [9], [10]. Our labelling approaches involve approximations that may introduce errors.

Future work should explore reformulating tasks to address class imbalance (binary classification, hierarchical approaches), alternative architectures (transformers, attention mechanisms), and expanding to the full VitalDB dataset. For endotype prediction, unsupervised discovery of phenotypes directly from non-invasive signals may be more tractable than predicting predefined categories. Ultimately, clinical utility depends on actionability and must be demonstrated through prospective trials showing improved outcomes [8], [9].

## V. CONCLUSION

We present a machine learning framework for predicting both severity and endotype of intraoperative hypotension using non-invasive measurements. Preliminary results on a subset of 180 VitalDB cases show moderate discriminative ability (severity AUROC 0.64, endotype AUROC 0.58) but poor classification accuracy (25% and 12%, respectively). Severe class imbalance and potential limitations of our labelling methodology, particularly for endotypes, significantly impact performance.

Despite these challenges, the moderate AUROC values suggest that discriminative features exist in non-invasive signals. Full-dataset training, comparison with MAP-based baselines, and task reformulation may improve performance. This work contributes to the field by employing rigorous patient-level cross-validation and transparent reporting of negative results. Final results incorporating the complete dataset and baseline comparisons will be presented at the conference.

## ACKNOWLEDGMENT

The authors would like to thank the VitalDB team for providing the publicly available dataset used in this study.

## REFERENCES

[1] M. Walsh u.a., Relationship between Intraoperative Mean Arterial Pressure and Clinical Outcomes after Noncardiac Surgery, Anesthesiology, Bd. 119, Nr. 3, S. 507–515, Sep. 2013, doi: 10.1097/ALN.0b013e3182a10e26.

[2] T. G. Monk u.a., Association between Intraoperative Hypotension and Hypertension and 30-day Postoperative Mortality in Noncardiac Surgery, Anesthesiology, Bd. 123, Nr. 2, S. 307–319, Aug. 2015, doi: 10.1097/ALN.0000000000000756.

[3] E. M. Wesselink, T. H. Kappen, H. M. Torn, A. J. C. Slooter, und W. A. van Klei, Intraoperative hypotension and the risk of postoperative adverse outcomes: a systematic review, Br J Anaesth, Bd. 121, Nr. 4, S. 706–721, Okt. 2018, doi: 10.1016/j.bja.2018.04.036.

[4] L. J. Wachtendorf u.a., Association Between Intraoperative Arterial Hypotension and Postoperative Delirium After Noncardiac Surgery: A Retrospective Multicenter Cohort Study, Anesth Analg, Bd. 134, Nr. 4, S. 822–833, Apr. 2022, doi:10.1213/ANE.0000000000005739.

[5] M. Wijnberge u.a., Association of intraoperative hypotension with postoperative morbidity and mortality: systematic review and meta-analysis, BJS Open, Bd. 5, Nr. 1, S. zraa018, Jan. 2021, doi: 10.1093/bjsopen/zraa018.

[6] M. M. Boltz, C. S. Hollenbeak, G. Ortenzi, und P. W. Dillon, Synergistic implications of multiple postoperative outcomes, Am J Med Qual, Bd. 27, Nr. 5, S. 383–390, 2012, doi: 10.1177/1062860611429612.

[7] F. Hatib et al., Machine-learning Algorithm to Predict Hypotension Based on High-fidelity Arterial Pressure Waveform Analysis, Anesthesiology, vol. 129, no. 4, pp. 663–674, Oct. 2018, doi: 10.1097/ALN.0000000000002300.

[8] J. Enevoldsen and S. T. Vistisen, Performance of the Hypotension Prediction Index May Be Overestimated Due to Selection Bias, Anesthesiology 137, no. 3: 283–89, https://doi.org/10.1097/ALN.0000000000004320.

[9] M. P. Mulder et al., Hypotension Prediction Index Is Equally Effective in Predicting Intraoperative Hypotension during Noncardiac Surgery Compared to a Mean Arterial Pressure Threshold: A Prospective Observational Study, Anesthesiology 141, no. 3: 453–62, https://doi.org/10.1097/ALN.0000000000004990.

[10] T. Tschoellitsch et al., Mean Arterial Pressure Is All You Need in a Machine Learning Model for Mean Arterial Pressure Prediction, European Journal of Anaesthesiology 42, no. 12: 1112–22, https://doi.org/10.1097/EJA.0000000000002238.

[11] E.M. Wesselink et al., Intraoperative Hypotension and the Risk of Postoperative Adverse Outcomes: A Systematic Review, British Journal of Anaesthesia 121, no. 4: 706–21, https://doi.org/10.1016/j.bja.2018.04.036.

[12] K. Kouz et al., Endotypes of Intraoperative Hypotension during Major Abdominal Surgery: A Retrospective Machine Learning Analysis of an Observational Cohort Study, British Journal of Anaesthesia 130, no. 3 (March 2023): 253–61, https://doi.org/10.1016/j.bja.2022.07.056.

[13] H.C. Lee, Y. Park, S.B. Yoon, S.M. Yang, D. Park, C.W. Jung, VitalDB, a high-fidelity multi-parameter vital signs database in surgical patients. Sci Data. 2022 Jun 8;9(1):279. DOI: 10.1038/s41597-022-01411-5.

 DOI 10.15203/99106-194-6-28

# Comparison of Noise-Reduction Methods for Unshielded Magnetorelaxometry for Magnetic Nanoparticles

Mina Naeimabadi, Aaron Jaufenthaler, Daniel Baumgarten

***Abstract*—The quantification of magnetic nanoparticles, which are used in different medical applications, is essential for reliable treatment planning and monitoring. To achieve this aim, accurate amplitude estimation is essential and can be obtained using magnetorelaxometry (MRX). However, performing MRX in an unshielded environment, which represents real clinical conditions, is challenging due to strong ambient interference. In this work, the gradient signal was derived from the two single channels of the sensor. Next, synthetic relaxation signals were generated by combining a noise-free simulated MRX relaxation curve with seven experimental noise signals. Three different filtering strategies and their combinations were then applied to the resulting signal. These included a forward-fitting step to suppress the power-line noise and its main harmonics, a moving-average step to reduce high-frequency sensor noise, and finally the Exponential–Sinusoidal Fitting step to remove the dominant sinusoidal tone of each pulse. Among all configurations, the forward-fitting stage resulted in the smallest amplitude error. Overall, the results of this study confirm that quantitative MRX reconstruction is feasible even under unshielded conditions and can enhance practical unshielded MRX measurements.**

***Index Terms*—magnetorelaxometry (MRX), magnetic nanoparticles, optically pumped magnetometer**

## I. Introduction

Magnetic nanoparticles (MNPs) play an important role in medical applications, including magnetic hyperthermia and drug delivery. Meanwhile, quantitative spatial information on the distribution of MNPs is crucial for monitoring and planning treatment. One promising method to obtain this information is magnetorelaxometry (MRX) imaging. In MRX, an DC magnetic field aligns the magnetic moments of the particles in the direction of the applied field; when the excitation field is switched off, the particle's magnetic moments relax back towards their easy-axis orientation. The relaxation amplitude is used for quantification and imaging of the particles [1], [2]. Whereas most studies in this area still rely on shielded environments, measuring MRX relaxation amplitudes in unshielded conditions represents a more realistic clinical setting. However, it also imposes the challenge of dealing with ambient magnetic interference. Even works performed in unshielded conditions have not been fully successful in addressing the main problem of environmental magnetic disturbances. Unlike continuous signals such as ECG, EEG, or MEG, noise removal in non-continuous MRX signals is challenging. Simple frequency-selective filtering methods, including notch filters, can introduce frequency-dependent phase distortions that affect the estimation of relaxation parameters [3]. For the first time, this work systematically compares different filtering methods and configurations in terms of preserving the MRX relaxation amplitude while removing noise as much as possible. Thereby, this approach contributes to bringing MRX imaging of magnetic nanoparticles closer to clinical application.

M. Naeimabadi, A. Jaufenthaler and D. Baumgarten are with the Biomedical engineering group, Department of Mechatronics, University of Innsbruck, Austria, e-mail: Mina.Naeimabadi@uibk.ac.at.

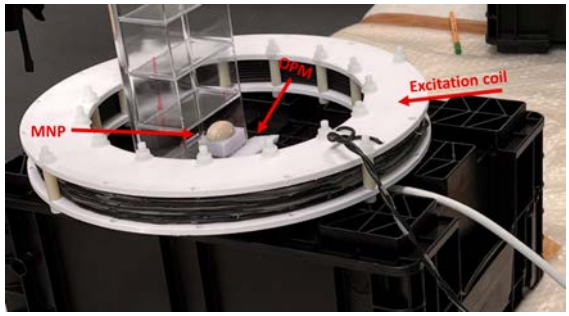


Fig. 1: Unshielded MRX setup including the pulsed OPM and excitation coil.

## II. Methods

The experiment was performed under unshielded laboratory conditions with the Earth's magnetic field as the background field. The MRX signals were measured using a highly sensitive two-channel total-field pulsed OPM (OMG – Optical Magnetic Gradiometer from Twinleaf LLC, NJ, USA) with a 2.3 cm channel separation and a sampling rate of 1 kHz. In addition, the setup included an excitation system consisting of a coil driven by a current source, as shown in Fig. 1. Two measurements, with and without nanoparticles, were recorded. A first-order gradiometer signal was obtained by subtracting the two OPM channels for each measurement. Each full measurement lasted 1.1 s, with a coil-on time of 100 ms and a relaxation/measurement time of 1 s. After identifying the first coil-on time, the full segment was extracted. Following the coil switch-off time, the first 10 ms of each cycle and the last 100 ms were discarded to account for coil ring-down and sensor dead time. As a result, the remaining signals were 890 ms long.

The available ten MRX pulses, over ten excitation cycles, were divided into two groups. For each group, the MRX pulses were averaged and the corresponding averaged empty pulses

 DOI 10.15203/99106-194-6-28

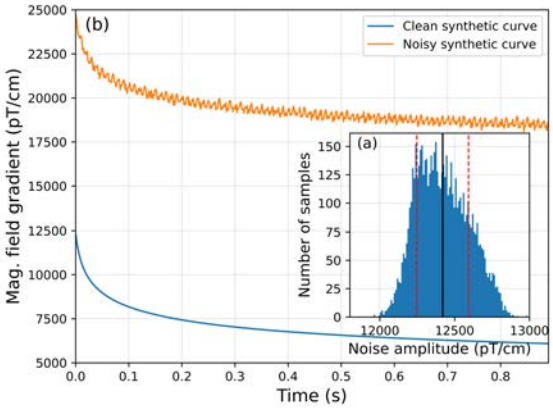


Fig. 2: Amplitude histogram of a representative noise-only signal with a mean value of $\mu = 12420.18$ pT/cm and a standard deviation of $\sigma = 172.64$ pT/cm (a). Synthetic MRX relaxation curve before and after adding real gradiometer noise (b).

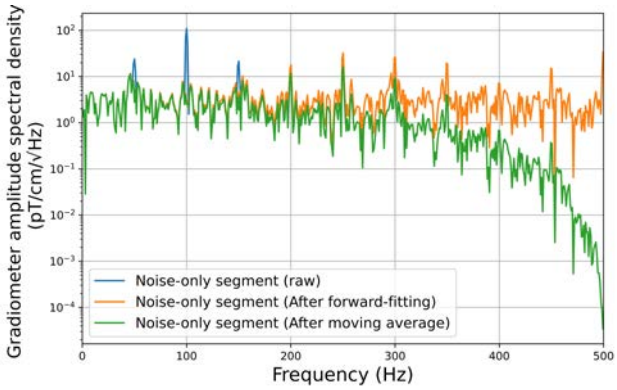


Fig. 3: Amplitude spectral density of the noise-only segment of the synthetic MRX relaxation signal. Before and after applying forward-fitting, and before and after applying the moving-average filter alone.

were then subtracted to remove common-mode and systematic noise.

The resulting MRX signals of each group were approximated as a sum of exponential functions using Orthogonal Matching Pursuit (OMP), which selected the four best-matching exponentials to represent the relaxation curves. From these, the fitted signal from the first group was then used as the clean synthetic curve, and two points at t1 = 0.1 s and t2 = 0.4 s were taken to subtract their amplitudes and to define the ground-truth amplitude. Next, each noise-only signal recorded separately under unshielded conditions were added to the clean synthetic curve from the second group to generate a set of synthetic signals with known relaxation amplitudes and real noise. The amplitude histogram of a representative noise-only recording is shown in Fig. 2 (a). Although the noise signals were already reduced by the internal gradiometer output of the OPM, they still contained dominant power-line interference and its harmonics. Both the clean and noise-induced MRX curves are illustrated in Fig. 2 (b).

The different filtering levels were then applied to this signal. The first filtering step was forward-fitting (FF). The noise-only part at the beginning of the generated noise-induced synthetic signal, with a length of 1.6 s, was used for forward-fitting. This duration was sufficient as it contained enough full cycles of the 50 Hz dominant noise and its harmonics. After removing the offset and detrending the noise-only segment, three dominant frequencies were selected around the main components at 50, 100, and 150 Hz. A multi-tone sinusoidal fit was constructed based on the chosen frequencies, and the fitted multi-tone signal was subtracted from the entire synthetic MRX signal. A moving average filter was then applied as the second cleaning step to reduce the 500 Hz internal sensor noise. For the final filtering step, each curve was modeled as a sum of four exponential functions and an offset. Once the exponential fit was found using OMP, the fitted curve was subtracted from the synthetic MRX curve and the amplitude spectral density (ASD) of the residual was used to identify the dominant remaining sinusoidal noise. The fitting was then repeated with a model combining four exponential functions, an offset, and a sinusoidal term based on the identified dominant frequency. In this step, only the sinusoidal component was removed to avoid distorting the relaxation amplitude. Different configurations of these steps were tested. The amplitude was recalculated using the OMP algorithm at the previously defined time points and the resulting error was computed for each configuration.

## III. Results

Following noise reduction through the gradiometer output of the two channels, all three filtering stages were individually applied to the synthetic signals. As a representative example, for one of the synthetic signals, in the forward-fitting step, three frequency tones were selected at 49.96, 99.98, and 149.96 Hz, close to the main components mentioned earlier. As shown in Fig. 3, this step strongly reduced the ASD at these tones by 85.7%, 97.1%, and 86.5%, respectively. The same figure illustrates the moving-average filter independently suppressing the broadband high-frequency sensor noise by 99.99%. The third step, Exponential–Sinusoidal Fitting (ESF), was applied to the synthetic MRX curve. In this case, the ASD at the most dominant sinusoidal residual frequency decreased from 138 pT/cm/$\sqrt{\text{Hz}}$ to 33 pT/cm/$\sqrt{\text{Hz}}$, corresponding to a 76% reduction down to the gradiometer's white noise floor.

Next, different configurations of these stages were evaluated. The amplitude of the filtered synthetic curve was calculated as explained before, and the percentage error relative to the ground-truth amplitude was computed. Among all combinations,applying the first step (FF) provided the best result. To give a clearer picture of the final outcome, Fig. 4 compares the signal with raw noise, the gradiometer result, and the final curve after applying the selected configuration. The second-best performance was achieved by the FF + ESF configuration, followed by MA + ESE. The corresponding amplitude errors are summarized in Table I as mean ± SD across seven noise scenarios.

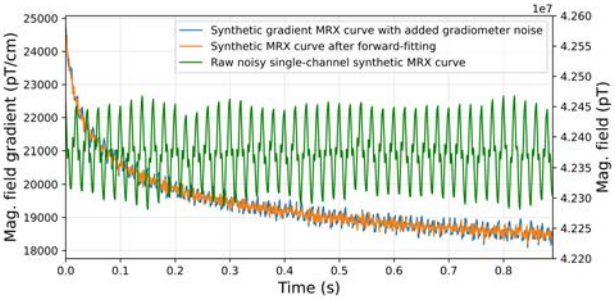


Fig. 4: Synthetic MRX signal at three stages: the raw noisy single-channel measurement, the MRX curve after adding gradiometer noise, and the curve approximated by forward-fitting.

TABLE I: Amplitude error (%) for each filtering configuration, reported as mean ± SD over seven synthetic signals. FF = forward-fitting; MA = moving-average; ESF = Exponential–Sinusoidal Fitting.

| Filtering Configuration | Amplitude error (%, mean ± SD) |
|---|---|
| FF | 0.30 ± 0.21 |
| FF + ESF | 0.40 ± 0.22 |
| MA + ESF | 0.91 ± 0.26 |
| MA | 1.06 ± 0.28 |
| FF + MA + ESF | 1.13 ± 0.22 |
| FF + MA | 1.15 ± 0.24 |
| ESF | 3.07 ± 0.29 |
| No filtering | 3.24 ± 0.26 |

## IV. Discussion

The main focus of each filtering step was on a different type of interference. Forward-fitting mainly removed the 50 Hz power-line noise and its first harmonics, whereas the moving-average filter targeted the broadband sensor noise around 500 Hz. In contrast, the Exponential–Sinusoidal Fitting operated on the relaxation curve to remove the most dominant remaining narrowband tone. Together, these results confirm that each filtering method has a different contribution to noise suppression.

Forward-fitting operates on the noise-only segment and reconstructs a fitted function from the three dominant frequency components. This method is used to remove dominant components before relaxation analysis, effectively suppressing ambient noise while preserving the overall signal shape. This explains its better performance in amplitude estimation compared to other strategies. The combination of forward-fitting and Exponential–Sinusoidal Fitting achieved the second-best performance. This behavior may occur because pre-filtering already modifies the signal in a way that influences the performance of the Exponential–Sinusoidal Fitting step, in terms of both the decay model and the identification of the most dominant residual frequency. Consequently, although the ESF approach operates directly on the relaxation curve and is designed to remove only the remaining residual tone, it is applied after other filtering and averaging–subtraction steps, making its performance sensitive to prior signal modifications. However, when combined with the moving-average filter, the results improved compared to using each method alone. Since the moving-average filter is less effective in removing lower-frequency noise, it introduces fewer pre-filtering effects, allowing Exponential–Sinusoidal Fitting to better identify and suppress the dominant residual frequency.

Other approaches ranked lower, likely due to overprocessing that altered the true amplitude.

Also the similar standard deviation across all filtering configurations indicates that the error spread is not caused by the filtering methods themselves. The different configurations change the mean amplitude error but do not significantly affect the overall deviation of the signals. Instead, the variability mainly comes from the data and is common to all cases.

For quantitative MRX, preserving the relaxation amplitude is more critical than minimizing the noise, since the amplitude is directly linked to the amount of particles. Although the best configuration performed well, a small amplitude error still remains because some residual noise cannot be fully removed and the fitted decay curve cannot perfectly match the true relaxation shape. Overall, the findings of this work demonstrate a promising step towards improving nanoparticle imaging in unshielded conditions.

This study, however, has some limitations. The dominant interference frequencies and sensor noise characteristics may differ across setups or environments. In the future, running the pipeline with MRX signals generated under different magnetic fields and excitation currents may lead to changes in the results. Further experiments will be needed to verify how stable these filtering steps are across different settings.

## V. Conclusion

In this work, different filtering strategies and their configurations were applied to a synthetic MRX relaxation curve generated from a real signal measured in an unshielded environment. Among all stages, the forward-fitting step showed the best performance in preserving the amplitude and producing the lowest error. These results suggest that quantitative estimation of magnetic nanoparticles using the MRX amplitude can potentially be performed even under unshielded conditions.

## VI. Acknowledgements

This research was funded in whole or in part by the Austrian Science Fund (FWF) 10.55776/PAT7880823.

## References

[1] P. Schier, M. Liebl, U. Steinhoff, F. Wiekhorst, D. Baumgarten, "Experimental Demonstration of Improved Magnetorelaxometry Imaging Performance Using Optimized Coil Configurations" *Medical Physics*, vol. 49, no. 5, pp. 3361–3374, 2022. doi:10.1002/mp.15594.

[2] A. Jaufenthaler, M. Sarzynska, F. Wiekhorst, L. Trahms, and J. Haueisen, "Pulsed Optically Pumped Magnetometers: Addressing Dead Time and Bandwidth for the Unshielded Magnetorelaxometry of Magnetic Nanoparticles," *Sensors*, vol. 21, no. 4, 1212, 2021. doi:10.3390/s21041212.

[3] D. Eberbeck, F. Wiekhorst, U. Steinhoff, and L. Trahms, "Aggregation behaviour of magnetic nanoparticle suspensions investigated by magnetorelaxometry," *Journal of Physics: Condensed Matter*, vol. 18, no. 38, pp. S2829–S2846, 2006.

ISBN 978-3-99106-194-6