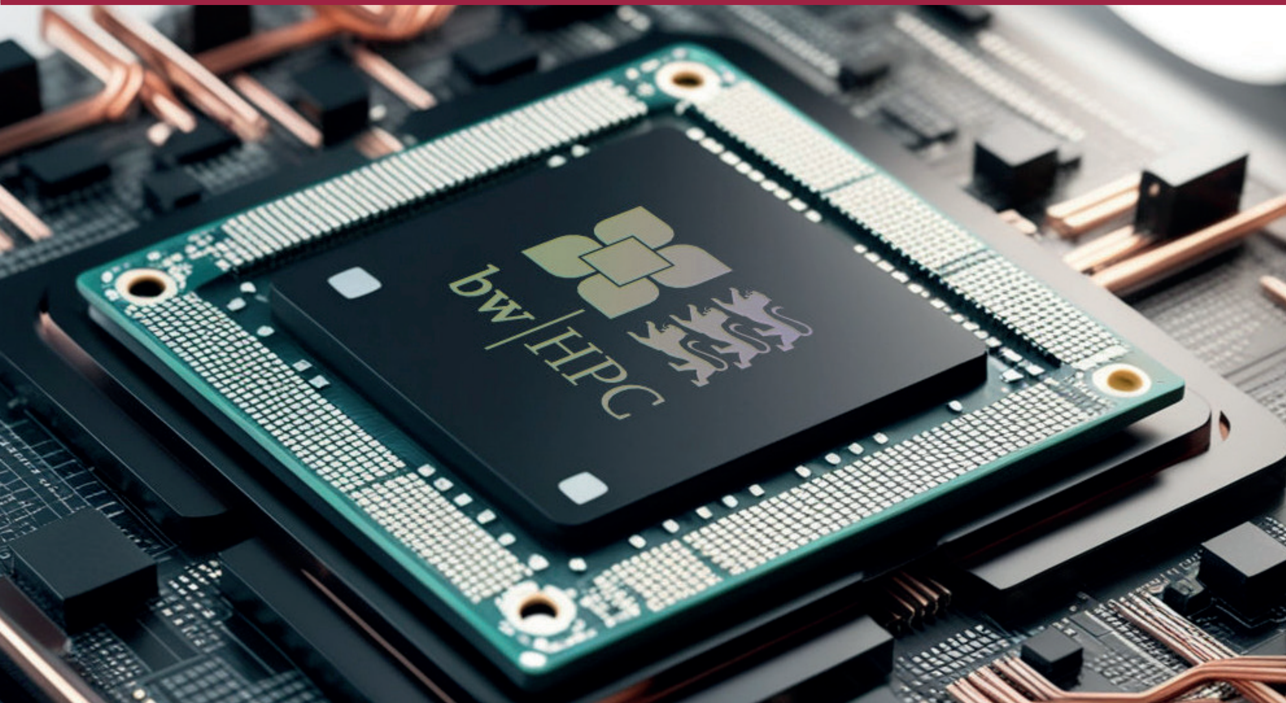
Jens Krüger
Suvasini Thangaraj
(Eds.)



Proceedings of the 11th bwHPC Symposium

High-Performance and Data-Intensive Computing in Baden-Württemberg

Tübingen, September 2025

Proceedings of the 11th

bwHPC Symposium

Proceedings of the 11th bwHPC Symposium

High-Performance and Data-Intensive Computing
in Baden-Württemberg

Tübingen, September 2025

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

TÜBINGEN
LIBRARY PUBLISHING

Bibliografische Information der Deutschen Nationalbibliothek
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie, detaillierte bibliografische Daten sind im Internet über http://dnb.d-nb.de abrufbar.

Die Online-Version dieser Publikation ist auf den Verlagswebseiten von Tübingen Library Publishing frei verfügbar (Open Access).
http://hdl.handle.net/10900/178650
http://nbn-resolving.org/urn:nbn:de:bsz:21-dspace-1786506
http://dx.doi.org/10.15496/publikation-119974

2026 Tübingen Library Publishing
Universitätsbibliothek Tübingen
Wilhelmstraße 32
72074 Tübingen
druckdienste@ub.uni-tuebingen.de
https://tlp.uni-tuebingen.de

ISBN (PDF): 978-3-98944-050-0

Umschlaggestaltung: Georgia Zechlin, Universität Tübingen
Coverfoto: Cover image created with leonardo.ai; design based on work by Bernd Walk

PREFACE

HIGH-PERFORMANCE COMPUTING IN BADEN-WÜRTTEMBERG: INSIGHTS FROM THE BWHPC SYMPOSIUM 2025 IN TÜBINGEN

In 2025, the bwHPC (Baden-Württemberg High Performance Computing) Symposium in Tübingen marked a pivotal moment in the evolution of high-performance computing (HPC) in Baden-Württemberg. The symposium, held at the renowned University of Tübingen, brought together researchers, technologists, operators, and policymakers to explore the frontiers of computational science, artificial intelligence, and technological innovation. As computational needs continue to grow across various fields—from climate modelling and genomics to machine learning and big data analytics—the bwHPC Symposium highlights the urgent need for collaboration, investment, and visionary strategies to ensure that Baden-Württemberg remains at the cutting edge of HPC advancements.

THE ROLE OF HIGH-PERFORMANCE COMPUTING IN MODERN RESEARCH

At its core, high-performance computing is the backbone of scientific discovery. In a world increasingly shaped by data, the need for vast computational power has never been more pronounced. Researchers and institutions are relying on HPC clusters to simulate complex systems, from the mechanisms of molecular function in drug discovery to the prediction of climate patterns. Tübingen, with its rich academic tradition, has become a central hub for these conversations, hosting experts who are redefining the boundaries of what is possible with advanced computing technology.

The bwHPC Symposium 2025 stands out not only for its technical discussions but also for its focus on the integration of HPC with other academic developments, such as artificial intelligence (AI) and research data management (RDM). These approaches are reshaping the landscape of computational science.

AI, for instance, is driving breakthroughs in predictive modelling, natural language processing, and automated systems. Yet such advancements require the computational horsepower that HPC platforms provide. For the scientific community, this means that understanding the synergies between HPC and AI is crucial for harnessing their full potential.

RDM and data-intensive computing have gained significant relevance over recent years. Within the data life cycle, processing environments such as HPC platforms play a crucial role.

The vast amounts of data generated in multiple research fields, including astrophysics, medical informatics and earth sciences, require powerful resources as provided through bwHPC.

A FEDERATED VISION FOR THE FUTURE OF HPC

The bwHPC Symposium 2025 in Tübingen serves as a critical juncture in Baden-Württemberg's pursuit of high-performance computing excellence. It represents an opportunity for the continent to solidify its leadership in computational science while fostering collaboration. With its focus on RDM, the integration of AI, sustainability, and scalable infrastructure, the symposium emphasized the importance of maintaining a long-term vision in the face of rapid technological evolution.

The Editors
Jens Krüger
Suvasini Thangaraj

For further information, please visit the symposium website at:
https://uni-tuebingen.de/de/274923
https://www.bwhpc.de/927.php

CHAIRS

Thomas Walter, University of Tübingen, Germany
Jens Krüger, University of Tübingen, Germany
Felix Bartusch, University of Tübingen, Germany
Stefan Behnle, University of Tübingen, Germany

LOCAL ORGANISERS

Felix Bartusch, University of Tübingen, Germany
Stefan Behnle, University of Tübingen, Germany
Raffel Doros, University of Tübingen, Germany
Ursula Eberhardt, University of Tübingen, Germany
Holger Gauza, University of Tübingen, Germany
Alexander Kirbis, University of Tübingen, Germany
Jens Krüger, University of Tübingen, Germany
Halima Saker, University of Tübingen, Germany
Zehra Hazal Sezer, University of Tübingen, Germany
Suvasini Thangaraj, University of Tübingen, Germany

PROGRAMME COMMITTEE

Martje Armbrecht, Karlsruhe Institute of Technology (KIT), Germany
Felix Bartusch, University of Tübingen, Germany
Martin Baumann, Heidelberg University, Germany
Stefan Behnle, University of Tübingen, Germany
Lucienne Dettki, University of Stuttgart, Germany
Rafael Doros, University of Tübingen, Germany
Ursula Eberhardt, University of Tübingen, Germany
Iven Fellhauer, Heidelberg University, Germany
Holger Gauza, University of Tübingen, Germany
Benjamin Gläßle, University of Tübingen, Germany
Uli Hahn, University of Tübingen, Germany

Michael Janczyk, University of Freiburg, Germany
Rainer Keller, Esslingen University of Applied Sciences, Germany
Alexander Kirbis, University of Tübingen, Germany
Philipp Köster, Esslingen University of Applied Sciences, Germany
Jens Krüger, University of Tübingen, Germany
Markus Quandt, University of Tübingen, Germany
Sabine Richling, Heidelberg University, Germany
Halima Saker, University of Tübingen, Germany
Jürgen Salk, University of Ulm, Germany
Hanna Schumacher, Heidelberg University, Germany
Herbert Steininger, University of Konstanz, Germany
Suvasini Thangaraj, University of Tübingen, Germany
Bernd Walk, University of Ulm, Germany
Bernd Wiebelt, University of Freiburg, Germany
Peter Weisbrod, Karlsruhe Institute of Technology (KIT), Germany
Brigitte Wellenkamp, University of Hohenheim, Germany
Hendrik Winkhardt, University of Mannheim, Germany
Lorenzo Zanon, University of Stuttgart, Germany

ABSTRACT

The 11th bwHPC Symposium showcases the breadth and depth of computational research enabled by high-performance computing across disciplines in Baden-Württemberg and beyond. Contributions span advances in Research Data Management (RDM), methodological innovation, domain sciences, and infrastructure development. Several works highlight the increasing integration of Research Data Management with HPC workflows, emphasizing FAIR principles, automation, and scalable data handling to support diverse scientific communities. In parallel, new HPC-enabled methods are presented for modeling complex electrochemical and quantum systems, including density functional theory and computational spectroscopy of electrochemical interfaces, and a novel diagrammatic high-temperature expansion approach to capture dynamical correlations in frustrated quantum magnets.

The proceedings feature significant strides in large-scale simulation and modeling, such as an accelerated Monte Carlo framework for long-axial field-of-view PET/CT systems, a massively parallel SYCL-based radiative transfer code for astrophysical applications, and refined CFD simulations to study airflow effects in agricultural machinery for insect conservation. Advances in astrophysics include N-body studies exploring the influence of galactic triaxiality on the evolution of triple supermassive black holes. In Earth system modeling, the introduction of PaleoPlaSim 1.0 demonstrates the potential of efficient, intermediate-complexity climate models to support large ensemble studies and perturbation-based calibration.

Contributions also highlight emerging security and AI-driven HPC challenges, with investigations into the detection of network information hiding methods at scale and reflections on the need for interactive, GPU-ready, interoperable HPC infrastructure to support modern AI-driven research workflows. Complementing these scientific advances, the bwRSE4HPC project illustrates how professional research software engineering practices enhance performance, sustainability, and collaborative development across scientific domains.

Together, these works underscore how HPC, combined with robust software engineering, modern infrastructure, and interdisciplinary methodologies, continues to drive scientific discovery, support new research paradigms, and enable the exploration of increasingly complex physical, biological, and computational systems.

On Sep 23rd, 2025, Attendees of 11th bwHPC Symposium were pictured in ZMBP, Tübingen

CONTENT

MARRYING TWO WORLDS: RESEARCH DATA MANAGEMENT AND SUPERCOMPUTING

STEPHAN HACHINGER[1*], DIETER KRANZLMÜLLER[1,2], LRZ FDM/RDM TEAM[1], INHPC-DE AP2 TEAM[1,3,4]

[1]LEIBNIZ SUPERCOMPUTING CENTRE (LRZ), BAVARIAN ACADEMY OF SCIENCES AND HUMANITIES, GARCHING B.M., GERMANY

[2]MNM-TEAM, LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN, MUNICH, GERMANY

[3]HIGH PERFORMANCE COMPUTING CENTRE STUTTGART (HLRS), UNIVERSITY OF STUTTGART, STUTTGART, GERMANY

[4]JÜLICH SUPERCOMPUTING CENTRE (JSC), FORSCHUNGSZENTRUM JÜLICH GMBH (FZJ), JÜLICH, GERMANY

[*]STEPHAN.HACHINGER@LRZ.DE

ABSTRACT

We reflect on a decade of work on Research Data Management (RDM) at Leibniz Supercomputing Centre (LRZ) and partner sites. Establishing a consistent RDM standard in High-Performance Computing (HPC) remains a challenge. We address this with a mix of a generic "least common denominator" approach and discipline-specific solutions, developed within special projects. A further focus of our work is the integration of RDM and data-driven, automatised workflows controlled with workflow engines inside or across systems, and on testing innovative backends for warm to cold data. We show work from our projects in different contexts, from local to EU scale, and emphasize the importance of properly managing expectations. From a FAIR (Findable, Accessible, Interoperable, Reusable) RDM solution, users do not only expect metadata and persistent identifier handling. Intuitively, they usually ask for performance, data safety, versatile analytics interfaces, full generality and ease of use at the same time. Projects, however, have to concentrate on a well-informed selection of these aspects to be successful.

Keywords: Research Data Management, High-Performance Computing (HPC), FAIR, Metadata, PIDs

1. INTRODUCTION

This contribution gives an overview of a decade of work by the Research Data Management (RDM) team at Leibniz Supercomputing Centre (LRZ, Garching b.M.), and concludes with a few lessons learned from our endeavours. Our work aims at enabling RDM methodologies in a computing-centre environment, and in particular in High-Performance Computing (HPC).

LRZ is the computing centre of the Munich universities and, at the same time, IT service provider for Bavarian universities and German High-Performance Computing (HPC) customers. As one of the three German top-tier HPC centres with HLRS (Stuttgart) and JSC (Jülich), it is part of the "Gauss Centre for Supercomputing" (GCS).

Huge datasets produced in supercomputing tend to be out of the scope of the RDM services that typical university libraries – as in Munich[i] – supply to their students and staff. Also, general-purpose or discipline-specific data repositories usually have problems accommodating such datasets towards the PB range. The LRZ RDM Team thus focuses on FAIR (i.e. Findable, Accessible, Interoperable, Reusable)[ii] RDM methods to fill the gap with a generic solution, and with strategies tailored to domain-specific and workflow settings. The goal of these efforts is to ensure that all researchers using the LRZ computing facilities find an appropriate service to publish data and make them FAIR.

2. THEORY AND METHODS

FAIR RDM has been conceptualised by Wilkinson et al.[ii] in 2016, in order to foster the re-use of scientific data and to address deficits in transparency and reproducibility. Their work defines measurable criteria on findability, accessibility, interoperability and reusability. In practice, the requirements[ii] focus on assigning globally-unique persistent identifiers (PIDs) to datasets and on generating, holding and publishing basic (e.g. DataCite[iii]) or extended metadata for each dataset. The metadata must be retrievable[ii] via the PIDs, which can be implemented via resolvers that redirect from the PID to a landing (web-)page with the metadata, e.g. https://doi.org as a resolver for Digital Object Identifiers (DOIs) as one kind of PIDs.

Practical RDM services for making data FAIR, thus make ample use of PIDs and metadata. They range from general-purpose or institute-specific platforms over domain-specific solutions to frameworks that target specific modes or methods of research and development (where the RDM is coupled to a source-code or computational workflow management or to electronic lab notebooks, for example). Below, we discuss our developments across this range.

3. RESULTS

A. GENERAL-PURPOSE FAIR SOLUTION FOR HPC IN GCS

Within the InHPC-DE project and previous work at LRZ, we designed a simple FAIR solution to make data FAIR that resides and remains on storage systems at the GCS centres (Figure 1). Building blocks for this have been developed and portal prototypes are running. Users are enabled to deposit DataCite[iii] metadata with their data. A python-based toolkit pushes the information into a web-based RDM framework (e.g. InvenioRDM[i]). Thus, a Persistent Identifier (PID) and a landing page are generated with instructions on how to obtain the actual data. Similar approaches[iv] are widely used and help bring data "out of the dark"[v].

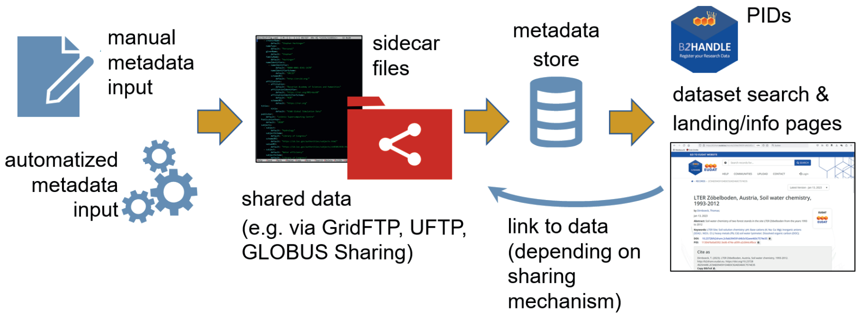


Figure 1: Data Publication Workflow in InHPC-DE

B. DOMAIN-SPECIFIC SOLUTIONS

At the moment, the FAIR HPC solution discussed above accepts generic DataCite[iii] and EUDAT[vi] metadata. To collaboratively work on discipline-specific methodical additions, we have entered collaborations such as DFG TRR356 PlantMicrobe[vii] and TRR419 SHARP (on learning via simulations). We are also working with consortia of the National Research Data Infrastructure (NFDI).

C. RDM IN COMPUTATIONAL WORKFLOWS

A further focus of our team has been the support of FAIR RDM in computational workflows. In particular, we have co-developed the data (and data-transfer) management in the LEXIS Platform, which facilitates workflows across European supercomputing centres[viii, ix]. We envisage enriching the data federation approach of LEXIS with diversified backends (object stores, databases) in the EXA4MIND[x] project on Extreme Data analytics.

4. CONCLUSION

The main lesson learned from all our projects probably concerns the importance of the "do one thing and do it well" principle. In RDM, a few things can be done at the same time, but caution is appropriate – as, e.g., storage performance engineering and metadata handling require vastly different expertise. Users or project partners usually demand FAIR capabilities, generality, high data-transfer rates, and deep analytics or semantic capabilities (e.g. easy inspection of data from defined geographical regions) from an RDM solution at the same time.

CHALLENGES FOR MODELING AND UNDERSTANDING ELECTROCHEMICAL INTERFACES: THE CASE OF INP

MATTHIAS M. MAY[1*], VIBHAV YADAV[1] AND HOLGER EUCHNER[1]

[1]INSTITUTE OF PHYSICAL AND THEORETICAL CHEMISTRY, UNIVERSITÄT TÜBINGEN, 72076 TÜBINGEN, GERMANY

[*]MATTHIAS.MAY@UNI-TUEBINGEN.DE

ABSTRACT

Electrochemical interfaces play an increasingly important role in a number of applications in the field of energy storage, but our atomistic understanding of these interfaces is, in most cases, still limited. Interface properties that need to be controlled range from (electro)chemical passivation to charge-selectivity by suitable electronic properties. Due to the fact that experiment struggles to convey information on the very interface through the — typically liquid — electrolyte, computational modeling of electrochemical interfaces is of great importance. However, modeling electrochemical interfaces can be challenging. Here, we discuss different approaches for addressing electrochemical systems by means of density functional theory. As for the case of exchange-correlation functionals, different levels of complexity with different computational cost and accuracy are available, corresponding to a Jacob's ladder of methods. This is exemplified in the case of InP (100) surfaces. Finally, we discuss how computational spectroscopy can be applied as the bridge between the computational model and experimental data.

Keywords: Electrochemistry, Density Functional Theory, Interfaces, Molecular Dynamics, Spectroscopy

1. INTRODUCTION

While the goal of achieving net-zero emissions by the middle of the century is largely agreed on, the necessary progress in emission reductions is still too slow. Cutting emissions is only feasible when the intermittent character of renewable energies can be compensated, which calls for the promotion of energy storage technologies. Apart from energy storage, carbon dioxide removal (CDR), often also referred to as negative emissions, will become necessary to compensate for remaining unavoidable emissions. In both fields, energy storage and CDR, electrochemical approaches are promising pathways that have the potential to significantly contribute to the envisaged net-zero emissions society[i]. However, developing efficient electrochemical devices calls for a detailed understanding of the underlying processes. This translates to the question of structure and dynamics at the electrochemical interface, which typically corresponds to a solid-liquid interface.

The microscopic behavior of this interface determines structural stability, but also properties such as potential distribution, surface state density or charge-transfer kinetics, which in turn are decisive for the overall performance of an electrochemical device. Hence, an atomistic understanding of the solid-liquid interphase is highly desirable. Yet, experimentally, transporting information about the interface through the liquid electrolyte is extremely challenging, such that, for an atomistic understanding, approaches combining experimental methods with electronic structure modeling are required[ii].

In analogy to the case of exchange-correlation functionals for density functional theory,[iii] a kind of *Jacob's ladder* of method can be introduced, meaning that the computational cost increases with the level of complexity (and accuracy) at which the electrochemical interface is investigated. In this work, we showcase different levels of approaches for the case of the semiconductor indium phosphide (InP), which is a prominent model system for photo-electrochemical water-splitting aiming at solar-driven hydrogen production.

2. RESULTS AND DISCUSSION

Following the above-introduced analogy to Jacob's ladder, the lowest level of complexity for modeling electrochemical interfaces corresponds to the investigation of surfaces in vacuum.

GEOMETRY OPTIMISATION IN VACUUM. A first understanding can already be gained by optimizing the geometry of plain surfaces in contact with constituents from the electrolyte – such as water molecules, in vacuum. To allow for a comparison between competing phases (of equivalent stoichiometry), surface energies are considered. By determining the most stable surface reconstructions with respect to the chemical potential of the respective elements, this approach can be extended to surfaces with different stoichiometry. For the case of InP, the surface phase diagram of the (100) surface and the respective surface reconstructions are depicted in Fig. 1a) and 1c). The latter has been obtained by applying the DFT code package CP2K[iv]. Exchange and correlation were accounted for by the PBE functional, whereas van der Waals interactions were represented via the Grimme-D3 correction[v, vi]. The phase diagram shows that distinct surfaces with increasing In-content are stabilized by an increasing indium chemical potential. For small values of the In chemical potential, the P-rich (2×2)-2D surface, a P-rich surface that is terminated by two additional P-dimers, is the most stable reconstruction, whereas the (2×4) mixed-dimer surface dominates at increased chemical potentials (see Fig. 1c). In the intermediate chemical potential range, the c(4×4) surface and the stepped α-(2×4) and β-(2×4) phases are stabilized.

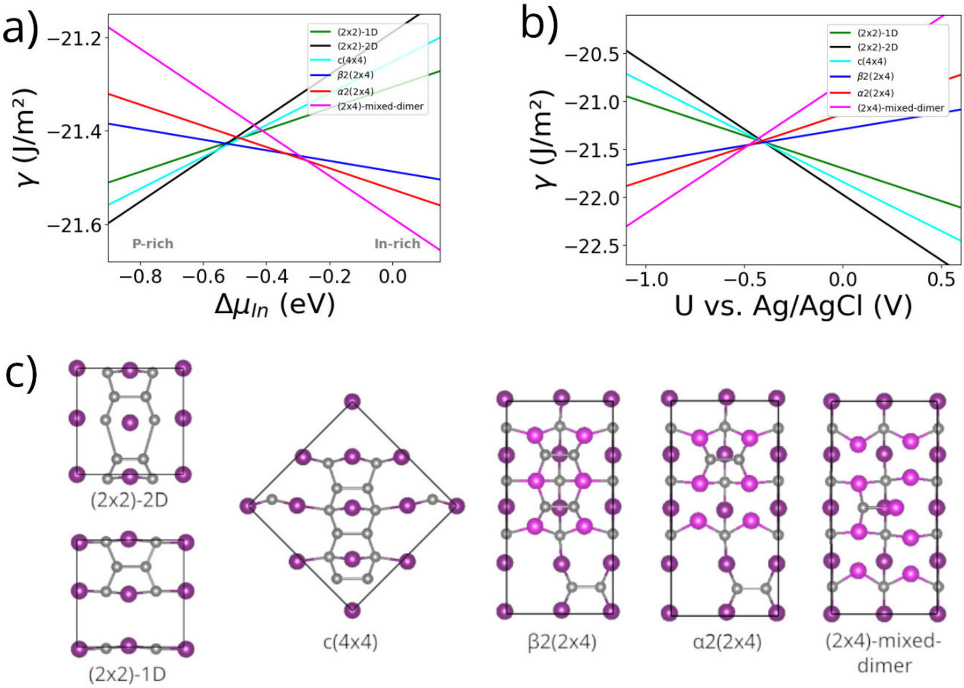


Figure 1: Phase diagram of the InP(100) surface as a function of a) the In chemical potential, b) the voltage as obtained via the computational hydrogen electrode approach. In c), the corresponding surfaces are depicted. Adapted from[viii].

This approach samples the structure-energy landscape, temperature effects and the electrolyte itself as well as fluctuations over time are, however, not considered.

COMPUTATIONAL HYDROGEN ELECTRODE. In the next step towards a more realistic modeling of electrochemical interfaces, such phase diagrams can be translated into the potential domain by applying the concept of the computational hydrogen electrode (CHE)[vii]. For the case of the InP(100) surface, this actually shows that the transition between the different stable phases falls in an extremely narrow potential window (see Fig. 1b). These phases have very different electrochemical properties, and it should be possible to switch between them by a variation of the applied potential in the electrolyte[viii]. This concept does, however, still neglect the actual surface chemistry, temperature effects, as well as dynamics and is therefore not able to account for the exact solid-liquid interactions at the interface. Nevertheless, it has proven to yield valuable insights on the impact of the electrochemical environment[viii].

CONSIDERING INITIAL CORROSION. At the next level of structural complexity, the interaction of the electrode surface with electrolyte species, such as e.g. hydrogen, oxygen or chlorine and the resulting structural changes, which one can consider as an initial step of corrosion, are investigated. This is exemplified for the case of oxygen adsorption on the InP(100) surface, where the corresponding phase diagram has been determined using the above-described computational setup (see Fig. 2). Here, a large fraction of the phase diagram is dominated by P-rich surfaces that are terminated by characteristic and rather stable P_xO_y polyphosphate motifs (see Fig. 2b). In particular, the β-(2×4)-based $\beta 2(P_2O_7)_3$, the $(2\times2)(P_2O_6)_2$ and the $(2\times2)(P_2O_7)_2$ surface, which are terminated by P_2O_6 and P_2O_7 units, cover a large part of the phase diagram[ix]. This approach will, depending on the reactivity and complexity of the surface, again increase the computational effort due to an increased size of the structural configuration space. Again, the translation to the electrochemical potential scale is then achieved by application of the CHE approach. While this approach tackles the previously mentioned disregard of the actual surface chemistry, it does so in a very static manner, still not considering the dynamic properties of the solid-liquid interface at finite temperature[x].

EVOLVING THE SYSTEM OVER TIME WITH MOLECULAR DYNAMICS. The next step is then to include the impact of the liquid phase in a molecular dynamics (MD) simulation, where finite temperatures are applied and the system evolved over time. This can be done by implicit or explicit solvent approaches, which either treat the solvent as a continuum or explicitly modeling also the solvent molecules on a DFT-level[xi]. Analysis of the MD trajectories allows for monitoring the overall dynamics of a system, whereas at the same time distinct configurations from snapshots of the trajectory can be studied in more detail. While the explicit, DFT-MD approach has the advantage of more realistically capturing chemistry, it is also more expensive, limiting the trajectory length typically to tens of picoseconds up to a few nanoseconds. This ansatz already captures many important aspects of the structurally variable solid-liquid interface, but neglects externally applied potentials and the resulting structural changes.

MOLECULAR DYNAMICS UNDER APPLIED POTENTIALS. In this step, applied potentials are introduced to the molecular dynamics simulation. The challenge here lies in how to approach electric fields under 3D-periodic boundary conditions, for which finite field molecular dynamics or the thermopotentiostat approach can be used[xii, xiii]. Depending on approach and implementation, the additional computational costs are moderate, which means that still, trajectory lengths of tens of picoseconds are computationally feasible. Yet still two major challenges remain: (1) As typical time spans for the electrochemical restructuring of surfaces range from seconds to hours[xiv], the currently available trajectories are many orders of magnitude shorter than effects observed in experiment. This means that only very fast effects at the solid-liquid interface can be captured, and the outcome of the simulation depends to a very large extent on the starting structure of the MD trajectory. (2) Currently affordable system sizes for long DFT-MD are in the order of hundreds to a few thousands of atoms. This enforces a high degree of ordering on the system and complex – amorphous, or 3d-structured systems cannot be reproduced.

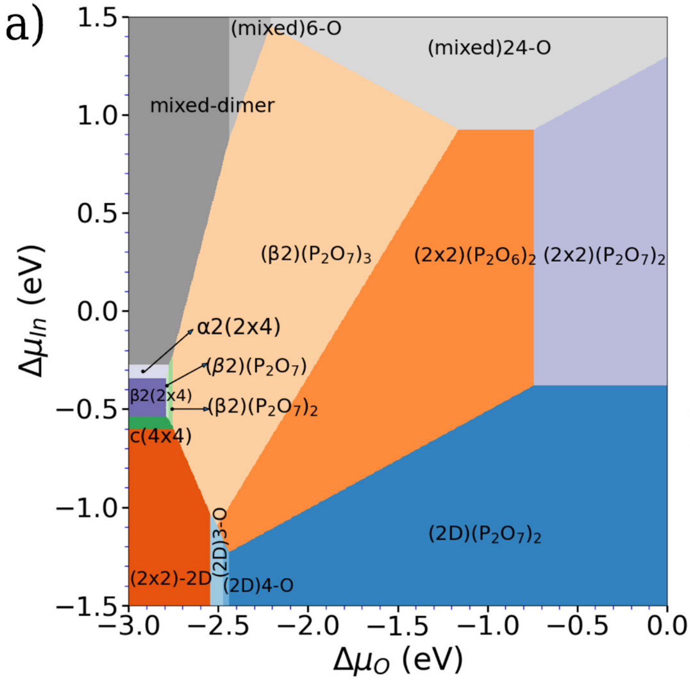
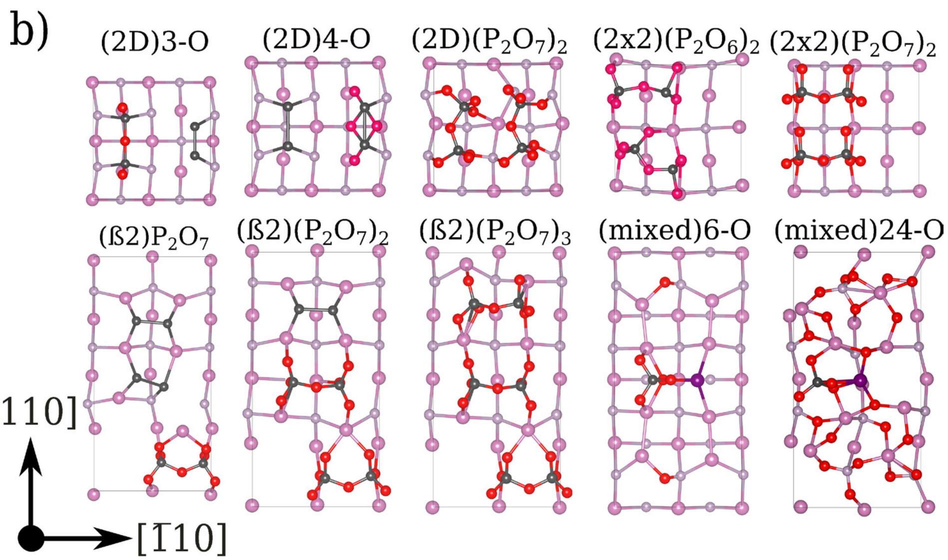


Figure 2a): Phase diagram of the InP(100) surface in contact with oxygen.

Figure 2b): Corresponding structures. Adapted from Reference xv



STRUCTURAL ANCHOR POINTS FROM EXPERIMENTAL INPUT. An approach that can at least partially alleviate the challenge of the short DFT-MD trajectories is to use structural information from experiment as starting points for simulations. Here, in the context of electrochemistry, recent studies have, for instance, applied a combination of operando reflection anisotropy spectroscopy and computational spectroscopy, which allows for indirectly probing the structural evolution of the oxygenated InP(100) surface[xv]. While computational spectroscopy helps to better understand the underlying atomistic structure of experimental data, experimental input can help to narrow down the number of structures to be considered as starting points for the modeling of electrochemical interfaces.

3. CONCLUSION

In this work, we have discussed methods on different levels of complexity for the modeling of electrochemical systems using the analogy to Jacob's ladder of accuracy with respect to exchange-correlation functionals. Computationally rather inexpensive approaches, in particular in the context of the computational hydrogen electrode, have successfully been applied in many cases. These approaches already allow us to gain general insights into the energetics of electrochemical systems. However, when dynamical aspects become important, DFT-MD based simulations are required to yield additional information. Yet as realizable trajectory lengths are still many orders of magnitude shorter than typical electrochemical timescales, input on structures from experiment is highly desirable to improve the realism of the simulation. Here, computational reflection anisotropy spectroscopy can serve as a bridge between experiment and electronic structure model. This has been exemplified for the case of the InP(100) surface, where different surface reconstructions under vacuum, the presence of oxygen and finally the structural dynamics have been investigated.

4. ACKNOWLEDGEMENT

The authors gratefully acknowledge funding from the German Federal Ministry of Research, Technology and Space (BMFTR), projects "NETPEC" (No. 01LS2103A) and "H2Demo" (03SF0619K), as well as DFG project number 434023472. The authors acknowledge support by the state of Baden-Württemberg through bwHPC and the German Research Foundation (DFG) through Grant No. INST 40/575-1 FUGG (JUSTUS 2 cluster).

5. REFERENCES

i M. M. May and K. Rehfeld, "Negative Emissions as the New Frontier of Photoelectrochemical CO2 Reduction", Adv. Energy Mater., vol. 12, p. 2103801, 2022. https://doi.org/10.1002/aenm.202103801

ii M. Guidat, M. Löw, M. Kölbach, J. Kim, and M. M. May, "Experimental and Computational Aspects of Electrochemical Reflection Anisotropy Spectroscopy: A Review", ChemElectroChem, vol. 10, no. 8, p. e202300027, 2023. https://doi.org/10.1002/celc.202300027

[iii] J. P. Perdew and K. Schmidt, ''Jacob's Ladder of Density Functional Approximations for the Exchange-Correlation Energy'', in AIP Conf. Proc., vol. 577, pp. 1–20, 2001. https://doi.org/10.1063/1.1390175

[iv] T. D. Kühne, et al., ''CP2K: An electronic structure and molecular dynamics software package - Quickstep: Efficient and accurate electronic structure calculations'', J. Chem. Phys., vol. 152, no. 19, p. 1941032020, 2020. https://doi.org/10.1063/5.0007045

[v] J. P. Perdew, K. Burke, and M. Ernzerhof, ''Generalized Gradient Approximation made simple'', Phys. Rev. Lett., vol. 77, no. 18, pp. 3865–3868, 1996. https://doi.org/10.1103/PhysRevLett.77.3865

[vi] S. Grimme, J. Antony, S. Ehrlich, and H. Krieg, ''A Consistent and Accurate ab initio Parametrization of Density Functional Dispersion Correction (DFT-D) for the 94 Elements H-Pu'', J. Chem. Phys., vol. 132, no. 15, p. 154104, 2010. https://doi.org/10.1063/1.3382344

[vii] J. K. Nørskov, J. Rossmeisl, A. Logadottir, L. Lindqvist, J. R. Kitchin, T. Bligaard, and H. Jónsson, ''Origin of the Overpotential for Oxygen Reduction at a Fuel-Cell Cathode'', J. Phys. Chem. B, vol. 108, no. 46, pp. 17 886–17 8892, 2004 https://doi.org/10.1021/jp047349j

[viii] H. Euchner, V. Yadav, and M. M. May, ''The InP(100) Surface Phase Diagram: From the Gas Phase to the Electrochemical Environment'', ACS Appl. Mater. Interfaces, vol. 17, pp. 8601–8609, 2025. https://doi.org/10.1021/acsami.4c20370

[ix] X. Zhang, T. Ogitsu, B. C. Wood, T. A. Pham, and S. Ptasinska, ''Oxidation induced polymerization of inp surface and implications for optoelectronic applications'', J. Phys. Chem. C, vol. 123, no. 51, pp. 30 893–30 902, 2019. https://doi.org/10.1021/acs.jpcc.9b07260

[x] O. M. Magnussen and A. Groß, ''Toward an atomic-scale understanding of electrochemical interface structure and dynamics'', J. Am. Chem. Soc., vol. 141, no. 12, pp. 4777–4790, 2019. https://doi.org/10.1021/jacs.8b13188

[xi] C. Zhang, T. Sayer, J. Hutter, and M. Sprik, ''Modelling electrochemical systems with finite field molecular dynamics'', J. Phys. Energy, vol. 2, no. 3, p. 032005, 2020. https://doi.org/10.1088/2515-7655/ab9d8c

[xii] F. Deißenbeck, C. Freysoldt, M. Todorova, J. Neugebauer, and S. Wippermann, ''Dielectric properties of nanoconfined water: A canonical thermopotentiostat approach'', Phys. Rev. Lett., vol. 126, no. 13, p. 136803, 2021. https://doi.org/10.1103/physrevlett.126.136803

[xiii] M. Löw, M. Guidat, J. Kim, and M. M. May, ''The Interfacial Structure of InP(100) in Contact with HCl and H2SO4 Studied by Reflection Anisotropy Spectroscopy'', RSC Adv., vol. 12, no. 50, pp. 32 756–32 764, 2022. https://doi.org/10.1039/d2ra05159a

[xiv] E. A. Schmitt, M. Guidat, M. Nusshör, A.-L. Renz, K. Möller, M. Flieg, D. Lörch, M. Kölbach, and M. M. May, ''Photoelectrochemical Schlenk cell functionalization of multi-junction water-splitting photoelectrodes'', Cell Reports Phys. Sci., vol. 4, no. 10, p. 101606, 2023. https://doi.org/10.1016/j.xcrp.2023.101606

[xv] V.Yadav, H.Euchner, and M. M.May, ''Surface Dynamics of Clean and Oxygenated InP(001) Surfaces in Contact with Water—Insights from Computational Spectroscopy''. ChemCatChem18, vol 18, no. 1, p. e01347, 2026. https://doi.org/10.1002/cctc.202501347

DETECTION OF NETWORK INFORMATION HIDING ARTIFACTS – A CASE FOR HPC?

STEFFEN WENDZEL[1*], JÖRG KELLER[2] AND SIMON VOLPERT[1]

[1]ULM UNIVERSITY, ULM, GERMANY

[2]FERNUNIVERSITÄT IN HAGEN, HAGEN, GERMANY

[*]STEFFEN.WENDZEL@UNI-ULM.DE

ABSTRACT

High-Performance Computing (HPC) is a barely considered domain when the criminal use of information hiding methods must be detected. We propose to investigate in more depth whether HPC environments might be a good choice to handle the growing amount of traffic to be analyzed for information-hiding-capable malware.

Keywords: Covert Channels, Steganography, Information Hiding, Malware, Anomaly Detection

1. INTRODUCTION

Information hiding aims to conceal secret information inside cover media, such as hiding cryptographic keys in network traffic characteristics. Malware increasingly applies these techniques to conceal secret data, for instance, when botmasters send commands to infected nodes. The detection of such malicious activity relies on identifying the subtle artifacts left in network traffic.

During the last ten years, the information hiding community has developed a unified taxonomy for hiding methods, referred to as hiding patterns[i]. While only a few major hiding patterns exist, these can be applied in arbitrary ways through so-called *pattern variation*. This results in tens of thousands of combinatorial options, creating a significant detection challenge. In large network environments, the amount of data to be investigated easily reaches petabytes, turning the detection of these numerous variations into a big data problem.

While it would be possible to analyze such volumes of network traffic for *specific* information hiding artifacts, their combinatorial possibilities render the detection of artifacts challenging for multiple reasons:

A. DISTRIBUTED ANALYTICAL CONTEXT

The analysis is fundamentally challenged by the fact that critical contextual information might be distributed across the entire petabyte-scale dataset. For a single flow, its packets may be spread across different recorded data chunks, making stateful pre-processing a communication-intensive task. Similarly, determining if a packet comprises an artifact requires context knowledge (i.e., a model of normal behavior) that can only be built by analyzing traffic patterns globally, not from isolated data blocks. Both issues hinder the effectiveness of simple, "embarrassingly parallel" analysis techniques.

B. QUALITY METRICS

A core goal of artifact detectors is to keep a minimum false-positive rate for a decent false-negative rate. While research regularly demonstrates that it can match these goals if their traffic is investigated for one particular hiding method, it is still considered an open problem to achieve good detection results for a large volume of traffic investigated jointly for multiple hiding patterns and their variations. Thus, one must resort to analyzing large traffic recordings for the high number of particular hiding methods coming from pattern variation.

Addressing these challenges at the petabyte-scale is hardly possible with commodity systems. HPC and related environments, however, provide the architectural capabilities to overcome this[ii]. The massive parallelism of HPC systems is essential for solving the combinatorial detection problem, while high-throughput parallel file systems and high-speed interconnects are critical for managing and preprocessing petabyte-scale network traces. While different flows can often be processed independently, the core challenge remains in managing the state of long-lasting flows distributed across many processing nodes. Communication-intensive frameworks like MPI are ideally suited to this task.

In this paper, we make the following contributions:

C1 We make the case for applying HPC to the problem of network covert channel detection, outlining how the challenges of combinatorial complexity and petabyte-scale data align with the capabilities of HPC environments.

C2 We present a concrete, scalable framework based on Dask for the parallel preprocessing of large-scale PCAP datasets, addressing the critical data ingestion bottleneck.

C3 We provide an initial performance evaluation of our framework on an HPC cluster, demonstrating its scalability and confirming the viability of our approach for handling large volumes of network traffic.

2. RELATED WORK

Multiple projects provide covert channel detection methods and/or traffic generation capabilities, e.g., BroCCaDe[iii] (based on the Bro IDS), CCGen[iv] (used for covert channel traffic generation) and pcapStego[v] (used to embed covert channels into PCAP files). However, none of these approaches considers HPC environments or highly parallel data processing. An exception is the WoDiCoF project[vi], which provides a parallel covert channel detection framework. The project has shown performance increase (speed-ups) on covert channel detection. However, WoDiCoF neglected the pre-processing issue as it took pre-processed flow data as given. Further, the project published their analysis for only one specific type of timing-based covert channel.

3. METHODOLOGY

To support our case for HPC in this domain (C1), we demonstrate a practical and scalable framework for parallel PCAP processing (C2), the computation graph of which is illustrated in Figure 1. This figure presents the high-level architecture and data flow within a distributed computing environment, specifically a Dask cluster[1] deployed on *bwUniCluster* 3.0.[2] It shows how distributed processing facilitates data ingestion and analysis, from initial data input to final result generation.

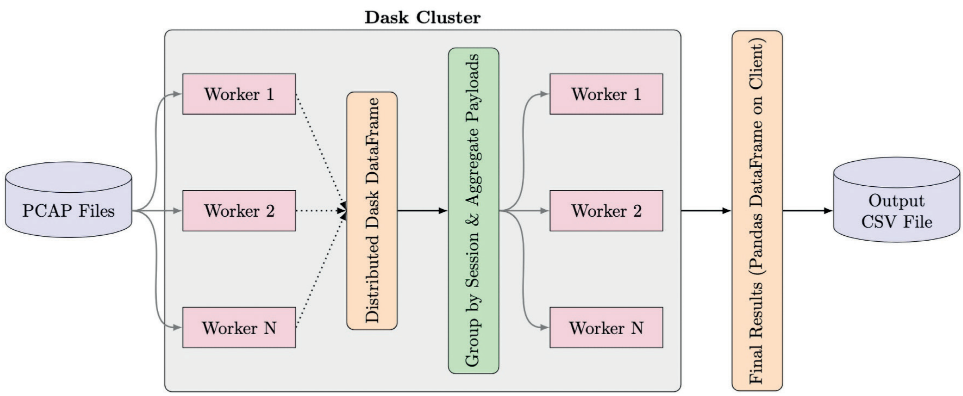


Figure 1: Our Dask-based Computation Graph for PCAP file analysis

The process begins with raw PCAP files, each containing approximately 1,000,000 packets, which are distributed among several worker nodes within the Dask cluster. Each worker independently reads and processes its assigned portion of the PCAP data, performing initial

[1] https://dask.org
[2] https://wiki.bwhpc.de/e/BwUniCluster3.0

parsing and feature extraction. This parallel ingestion leads to the creation of a distributed Dask DataFrame, which is a lazy, parallel DataFrame composed of many Pandas DataFrames. This distributed data structure allows for computations to be performed in parallel across the cluster, reducing the analysis time for large PCAP datasets. The inherent parallelism and out-of-core computing capabilities of such distributed frameworks, exemplified by Dask, are important for handling gigabytes or even terabytes of PCAP data that would be challenging on a single machine, thus supporting scalability. Subsequent operations, such as grouping by session and aggregating payloads, are performed effectively across the distributed DataFrame, utilizing the combined computational power of the worker nodes. This distributed approach allows the system to scale horizontally by adding more workers to the cluster as the data volume increases. Finally, the processed and aggregated data are collected back to a single Pandas DataFrame on the client, which can then be easily saved to an output CSV file for further analysis or reporting, such as generating statistical summaries.

To validate our framework, we performed a scalability evaluation (C3) by running experiments on clusters of 2, 4, and 8 nodes. Each node in the cluster was configured with 8 CPU cores and 32 GB of memory. The experiment involved processing 10, 20, 30, and 40 PCAP files, and measuring the time taken for the analysis to complete.

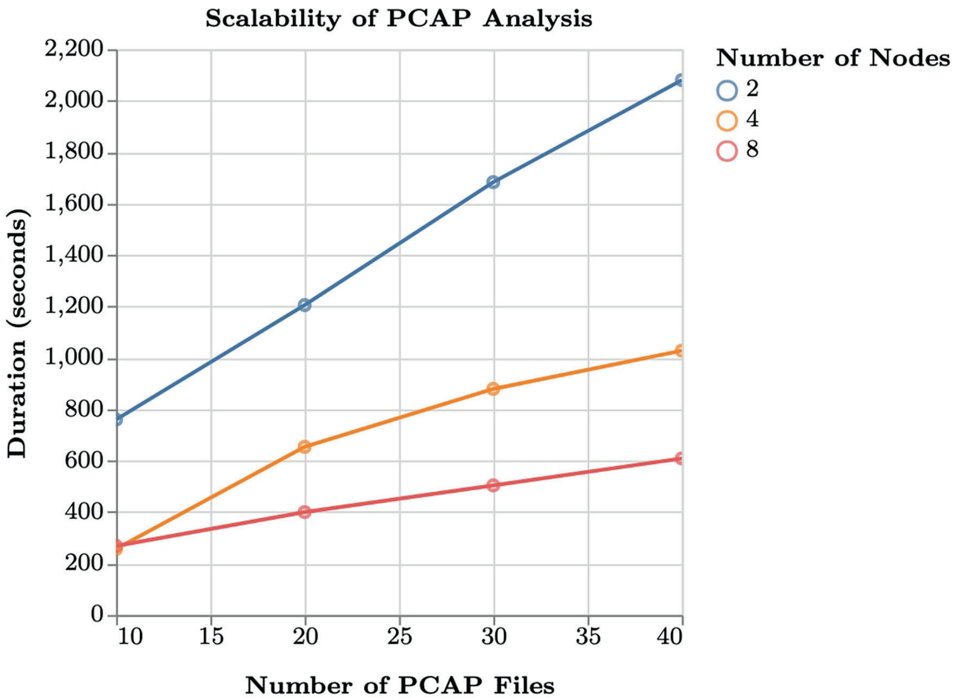


Figure 2: Scalability of PCAP Analysis with Different Cluster Sizes

The results, presented in Figure 2, show a clear trend: as the number of nodes in the cluster increases, the time required to process the PCAP files decreases significantly. This behavior indicates that the distributed processing approach scales well with the number of computational resources. The lines for each node configuration show a near-linear relationship between the number of files and the processing duration. This is a positive indicator of scalability, as it suggests that the system can handle increasing amounts of data by adding more nodes, with a predictable increase in performance. The observed results confirm that leveraging a high-performance computing cluster is a viable and efficient strategy for analyzing large volumes of PCAP data.

4. CONCLUSION AND FUTURE WORK

We addressed whether covert channel detection can be considered a relevant case for HPC environments. We believe that this is the case due to the combinatorial possibilities of having thousands of hiding methods to embed secret data in network flows while the volume of data exchanged on the Internet grows massively. To address the analysis of such large-scale traffic data for information hiding artifacts, HPC environments are a promising resource. Given the promising performance of our initial experiment, future work will focus on coupling our framework with a range of HPC-ready covert channel detection modules. This will involve implementing several more detection methods to create a complete, HPC-ready process from preprocessing to detection.

5. REFERENCES

[i] S. Wendzel, L. Caviglione, W. Mazurczyk, A. Mileva, J. Dittmann, C. Krätzer, K. Lamshöft, *et al.*, "A generic taxonomy for steganography methods", Computing Surveys, vol. 57, no. 9, pp. 1–37, 2025. https://doi.org/10.1145/3729165

[ii] D. S. Terzi, R. Terzi, and S. Sagiroglu, "Big data analytics for network anomaly detection from netflow data", in 2017 Int. Conference on Computer Science and Engineering, pp. 592–597, 2017. https://doi.org/10.1109/UBMK.2017.8093473

[iii] H. Gunadi and S. Zander, "Bro covert channel detection (BroCCaDe) framework: scope and back-ground", tech. rep., Murdoch University, 2017.

[iv] F. Iglesias, F. Meghdouri, R. Annessi, and T. Zseby, "CCgen: Injecting covert channels into network traffic", Security and Communication Networks, vol. 2022, 2022. https://doi.org/10.1155/2022/2254959

[v] M. Zuppelli and L. Caviglione, "pcapStego: A tool for generating traffic traces for experimenting with network covert channels", in Proc. Int. Conf. Availability, Reliability and Security, pp. 1–8, 2021. https://doi.org/10.1145/3465481.3470067

[vi] R. Keidel, S. Wendzel, S. Zillien, E. S. Conner, and G. Haas, "WoDiCoF – A testbed for the evaluation of (parallel) covert channel detection algorithms", Journal of Universal Computer Science, vol. 24, no. 5, pp. 556–576, 2018. https://doi.org/10.3217/jucs-024-05-0556

DYNAMIC CORRELATIONS OF FRUSTRATED QUANTUM SPINS FROM HIGH-TEMPERATURE EXPANSION

BJÖRN SBIERSKI*

INSTITUT FÜR THEORETISCHE PHYSIK, UNIVERSITÄT TÜBINGEN, 72076 TÜBINGEN, GERMANY

*BJOERN.SBIERSKI@UNI-TUEBINGEN.DE

ABSTRACT

Models of interacting spins are paradigmatic descriptions of highly correlated and entangled quantum systems that can harbor exotic phases of matter like highly entangled quantum spin liquids. The latter are potentially relevant technologically to facilitate quantum information processing or high-temperature superconductivity. The most interesting spin models are characterized by frustrated interactions, which means that quantum fluctuations are drastically enhanced. However, the theoretical description of frustrated quantum magnets remains difficult, especially in dimensions two and three, at non-zero temperature and in particular for dynamical properties. Here, we summarize key aspects of a novel diagrammatic approach to the dynamic spin structure factor that we developed recently.

Keywords: Quantum Magnetism, Series Expansions, Quantum Dynamics

1. INTRODUCTION

For the theoretical analysis of frustrated spin models, the technique of high-temperature series expansion is a well-established tool conventionally addressing thermodynamic quantities and equal-time spin correlations, see Ref.[i] for a review. In recent work (Refs.[ii,iii]) we extended the applicability of this expansion to the dynamic (time-dependent) spin correlation functions, $\langle S^z_i(t) S^z_{i'} \rangle$. After Fourier transforming to frequency and momentum space, these quantities characterize the nature of excitations in the system and help to distinguish exotic from more mundane phases of matter.

2. THEORY AND METHODS

We focus on Heisenberg models $H=\Sigma_{i<i'} J_{ii'} (S^x_i S^x_{i'} + S^y_i S^y_{i'} + S^z_i S^z_{i'})$ with a single coupling constant J and spin lengths S=1/2 and S=1. As a first step, we develop an essentially analytic formalism to compute expansion coefficients for the imaginary frequency Matsubara spin correlators. We consider its expansion coefficients up to 12th order in x=J/T on lattice snippets, which we call graphs. Up to this order, there are about 10^6 graphs embeddable in arbitrary lattices, and we

evaluate the Matsubara correlator exactly on all of them using HPC systems. For accuracy, it is important to use integer-based arithmetic via rational numbers.

The first result at zero Matsubara frequency is the static susceptibilities for arbitrary site-pairs or wavevectors. Moreover, the analytic frequency dependence in the expansion allows for stable analytic continuation to the real-frequency dynamic structure factor using the technique of continued fraction expansion of the spectral function.

3. RESULTS AND DISCUSSION

Experimentally, the dynamic structure factor is routinely measured in inelastic neutron scattering. We compare our results for an anti-ferromagnetic S=1 nearest-neighbor Heisenberg model on the heavily frustrated three-dimensional pyrochlore lattice (corner-sharing tetrahedra) to the neutron scattering cross-section on the material $NaCaNi_2F_7$ which is believed to realize this model approximately. The comparison in Fig. 1 indeed shows fair agreement.

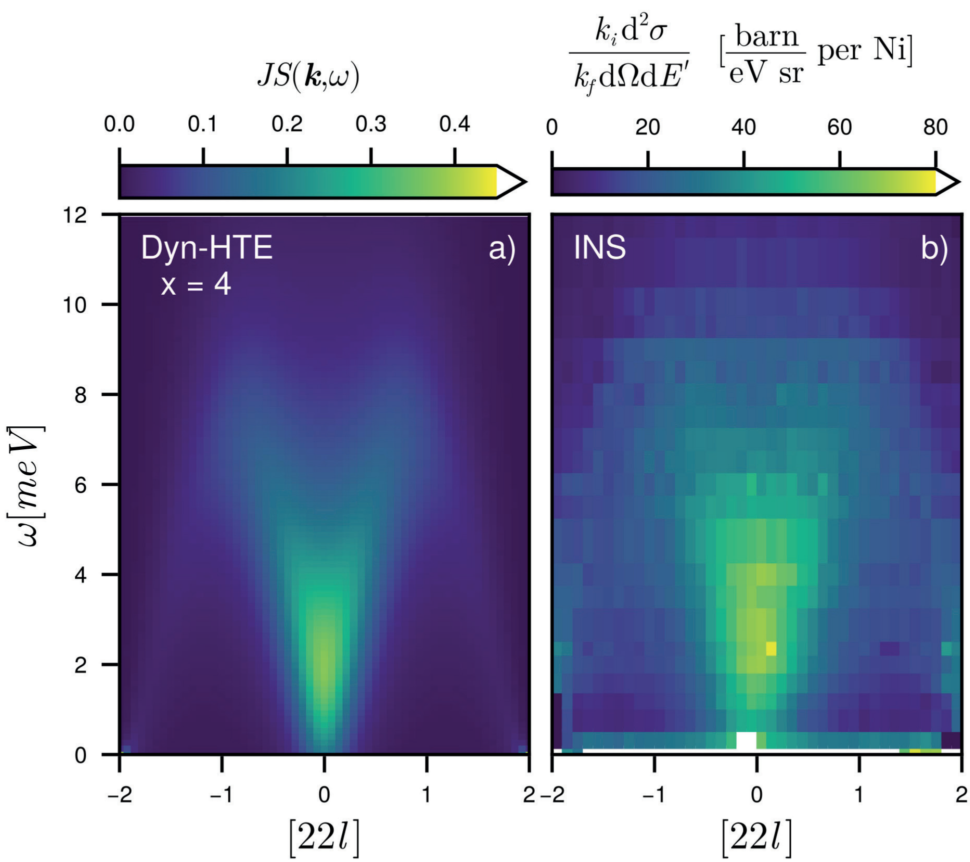


Figure 1: (a) Dynamic structure factor for the pyrochlore lattice S=1 Heisenberg anti-ferromagnet for k = (2,2,l), J=2.4meV at J/T=x=4 from Dyn-HTE. (b) Experimental INS data for $NaCaNi_2F_7$ at x=15, from Ref.[iv]

4. CONCLUSION

Dyn-HTE is an emerging new method for the numerical calculation of the finite-temperature spin dynamics for high-dimensional frustrated magnets. Due to its graph-based nature, pre-computed graph evaluations are easily embedded in arbitrary lattices. Future extensions include the calculation of higher-order dynamic correlators, quantum Fisher information as an entanglement witness or the generalization to beyond-Heisenberg Hamiltonians with reduced symmetry.

5. ACKNOWLEDGEMENT

The author also acknowledges support by the state of Baden-Württemberg through bwHPC and the German Research Foundation (DFG) through Grant No. INST 40/575-1 FUGG (JUSTUS 2 cluster).

6. REFERENCES

[i] J. Oitmaa, C. Hamer, and W. Zheng, Series Expansion Methods for Strongly Interacting Lattice Models, Cambridge University Press, Cambridge, 2010. https://doi.org/10.1017/CBO9780511584398

[ii] R. Burkard, B. Schneider, B. Sbierski, High-temperature series expansion of the dynamic Matsubara spin correlator, Phys. Rev. B, 113.075102, 2026. doi: https://doi.org/10.1103/1l92-z6qd

[iii] R. Burkard, B. Schneider, B. Sbierski, Dynamic correlations of frustrated quantum spins from high-temperature expansion, Phys. Rev. Lett., 136.056501, 2026. doi: https://doi.org/10.1103/jtjk-x2lw

[iv] K. W. Plumb, H. J. Changlani, A. Scheie, S. Zhang, J. W. Krizan, J. A. Rodriguez-Rivera, Y. Qiu, B. Winn, R. J. Cava, and C. L. Broholm, Continuum of quantum fluctuations in a three-dimensional S = 1 Heisenberg magnet, Nature Physics, vol. 15, p. 54, 2019. https://doi.org/10.1038/s41567-018-0317-3

ACCELERATED MONTE CARLO SIMULATION FRAMEWORK FOR PATIENT MOTION CORRECTION WITH A LONG AXIAL FIELD OF VIEW POSITRON EMISSION TOMOGRAPHY SCANNER

WENHONG LAN[1]*, CHRISTIAN POMMRANZ[2], STEFAN WEIGEL[1], PIA M LINDER[1], ELNAZ OLYAEI[1,3], CHRISTIAN LA FOUGÈRE[1,4] AND FABIAN P SCHMIDT[1,3]

[1] DEPARTMENT OF NUCLEAR MEDICINE AND CLINICAL MOLECULAR IMAGING, UNIVERSITY HOSPITAL TÜBINGEN, TÜBINGEN, GERMANY

[2] INSTITUTE FOR ASTRONOMY AND ASTROPHYSICS, UNIVERSITY OF TÜBINGEN, TÜBINGEN, GERMANY

[3] WERNER SIEMENS IMAGING CENTER, DEPARTMENT OF PRECLINICAL IMAGING AND RADIOPHARMACY, UNIVERSITY OF TÜBINGEN, TÜBINGEN, GERMANY

[4] CLUSTER OF EXCELLENCE IFIT (EXC 2180) "IMAGE GUIDED AND FUNCTIONALLY INSTRUCTED TUMOR THERAPIES", UNIVERSITY OF TÜBINGEN, TÜBINGEN, GERMANY

*WENHONG.LAN@MED.UNI-TUEBINGEN.DE

ABSTRACT

Background: Long Axial Field-of-View (LAFOV) PET scanners provide unprecedented sensitivity and spatio-temporal resolution, enabling whole-body dynamic imaging and quantitative studies. However, evaluating performance aspects such as patient motion correction is challenging in clinical settings, as true ground-truth data cannot be acquired from patients. To address this, we developed a comprehensive Monte Carlo simulation framework for the Biograph Vision Quadra LAFOV PET/CT system, integrating anatomically realistic, motion-resolved digital phantoms.

Methods: The framework, based on OpenGATE, reproduces detailed detector geometry, event physics, and acquisition processes. In this study, we implemented a high-performance computing-based workflow on the bwHPC-NEMO2 cluster to overcome two key bottlenecks: (1) *the large number of simulated particles required for LAFOV systems* and (2) *the generation and handling of high-resolution 4D phantoms*. Respiratory motion from real patients was incorporated into 4D XCAT phantoms, and simulations were parallelized using job arrays. The pipeline included data conversion via root2lm and image reconstruction using the vendor's software, consistent with clinical practice.

Results: A benchmark using 132 MBq of F-18, 300 s scan duration (1000 frames, 0.3 s/frame), and 14 lesions demonstrated a total runtime of 21 hours on the cluster—29 times faster than a local workstation (>1 month). Furthermore, the HPC environment enabled 1.5 mm³ voxel resolution modeling, substantially improving anatomical accuracy.

Conclusion: The results validate that integrating LAFOV PET simulations with HPC infrastructure enables high-fidelity, time-efficient modeling of realistic patient conditions, offering a robust foundation for future developments in motion correction and quantitative PET imaging research.

Keywords: Monte Carlo Simulation, LAFOV PET/CT, High-Performance Computing.

1. INTRODUCTION

Long Axial Field-of-View (LAFOV) Positron Emission Tomography (PET) scanners with their high sensitivity combined with high spatio-temporal resolution hold an unprecedented potential for both clinical and research applications such as dosimetry optimization[i] and holistic total-body PET dynamic studies[ii]. However, certain critical information such as assessing the performance of patient motion correction also for these scanners cannot be directly obtained from real patient scans. In this context, accurate Monte Carlo (MC) simulations using anatomically realistic digital phantoms are essential to generate ground-truth data. A Geant4 Application for Emission Tomography (GATE) based simulation[iii] and reconstruction framework for the Biograph Vision Quadra LAFOV PET/CT scanner (Siemens Healthineers, TN, USA) was developed by our group[iv], capable of mimicking the real event physics, data acquisition and reconstruction processes together with a realistic phantom mimicking a patient with realistic respiration motion[v]. But two major computational bottlenecks remain. First, the large number of billions of particles required to simulate a realistic scan with the LAFOV PET system. Second, the generation and handling of high-resolution, 4D digital phantoms, to represent both anatomical details and small lesions with high resolution. These simulations cannot be performed on local workstations due to excessive computing and I/O demands. In this study, we developed an HPC-based workflow to accelerate an end-to-end MC simulation framework with the bwHPC-NEMO2 cluster and offer a benchmark test to show the improvements and benefits of HPC for MC simulation at LAFOV PET system.

2. THEORY AND METHODS

The Monte Carlo simulations are CPU-based, and long scan durations can be split into shorter frames by job array, offering the opportunity to be accelerated by parallel execution on the cluster. Patient-like 4D extended cardiac-torso (XCAT) phantoms[vi] were used in the framework, and lesions and anatomical structures were defined and generated. Respiratory motion was based on real patient-specific respiration and subsequently merged the lesions and the anatomical structures on the cluster. The blank spaces in the voxel matrices were cropped to minimize the RAM demand during simulation. The resulting voxel matrices were formatted to float32 digit precision to minimize the storage demand and converted to interfile data format aiming to be compatible with OpenGATE simulation framework, incorporating a detailed digital replica of the Quadra system. The MC simulation was executed on bwHPC NEMO2 using job arrays, with each task simulating a short frame in range of 0.3 seconds. Output data was post-processed by an investigational software *root2lm* (Siemens Healthineers, TN, USA), which processes and converts the single events stored in the ROOT files (output of the MC simulation) to coincidence event binary data (listmode) compatible with image reconstruction software. Subsequently image reconstruction is performed for the listmode data using the vendor's image reconstruction investigational prototype *e7tools* (Siemens Healthineers, TN, USA), which is identical to the clinical reconstruction pipeline. The simulation and image reconstruction framework and workflow is illustrated in Figure 1.

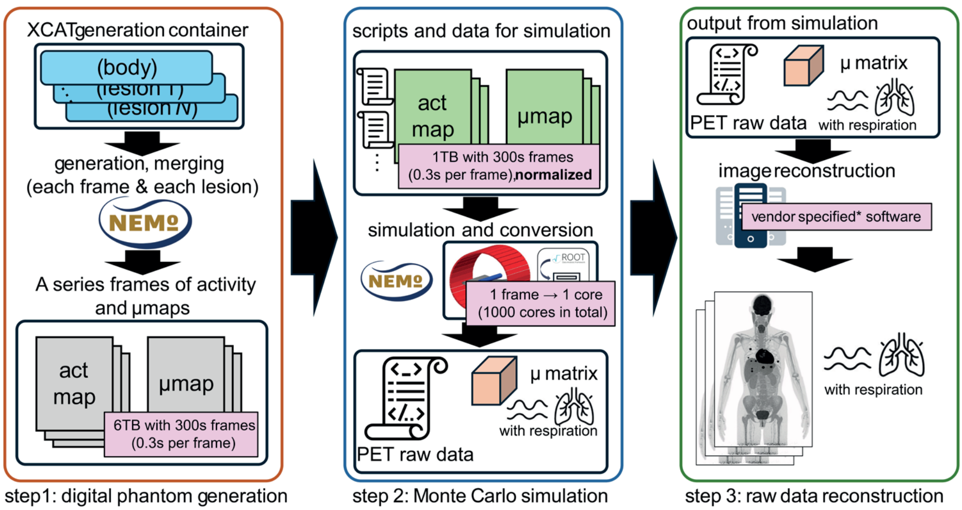


Figure 1: Workflow of the simulation and image reconstruction framework of the Biograph Vision Quadra *LAFOV PET scanner* with realistic patient phantom modelling respiratory motion using *bwHPC-NEMO2 cluster*

A total of 14 lesions with typical locations and standard uptake values (SUV) were modeled based on a cohort of 20 oncological subjects. A benchmark test for computation time with the

XCAT phantom with the 14 lesions and with real patient respiration (132 MBq F-18) for a 300 s scan (1000 frames,0.3 s per frame) was conducted on the cluster and a local workstation (Intel Core® i9-14900kf, 128 GB DDR4 RAM and 8 TB HDD). For preliminary assessment of the potential to use this framework for validation of motion correction algorithms, the motion correction was performed via the software OncoFreezeAI (Siemens Healthineers, TN, USA[vii]).

3. RESULTS AND DISCUSSION

Leveraging NVMe storage and 100 GbE interconnects, the total runtime was 21 hours using parallel execution on a cluster, 29 times faster than a local workstation (>1 month). Moreover, the ample DDR5 memory enabled high-resolution voxel modeling (1.5 ×1.5×1.5 mm), increasing simulation realism. In contrast, the voxel size of the XCAT phantom by local workstation usually cannot be smaller than 3.1 ×3.1×3.1 mm, due to the restriction of the maximum RAM during the simulation. The comparison of each step (except reconstruction) is shown in Table 1. The images derived from the exemplary simulation datasets of no motion implemented, motion implemented, and motion correction are shown in Figure 2 (B) shows the severe image blurring introduced by respiration motion, in comparison to the reference (Figure 2 (A)). After implementing the motion correction, the motion artefact was partly compensated, and the image blurring was reduced (Figure 2 (C)).

	XCAT GENERATION	MC SIMULATION	POSTPROCESSING
local workstation	73 h (3.1×3.1×3.1 mm)	21 d	1.5 d
NEMO	5 h (1.5×1.5×1.5 mm)	15 h	40 mins

Table 1: Time consumption of each step between the cluster and the local workstation

Figure 2: (A) The XCAT phantom with 14 defined lesions at lung and liver areas. **(B)** The 14-lesion XCAT phantom with patient extracted respiration motion. **(C)** The motion correction was implemented into the 14-lesion XCAT phantom with motions.

4. CONCLUSION

The integration of our simulation framework with the bwHPC-NEMO cluster enables high-fidelity, time-efficient MC simulations for a LAFOV PET system. This scalable platform provides a crucial foundation for future research in motion correction and image reconstruction under clinically realistic conditions

5. ACKNOWLEDGEMENT

The authors would like to thank the state of Baden-Württemberg for the support of the bwHPC and the German Research Foundation (DFG) for funding under the reference „INST 39/1232-1 FUGG“ (bwForCluster NEMO 2).

6. REFERENCES

i P. M. Linder et al., "Optimization of Y-90 Radioembolization Imaging for Post-Treatment Dosimetry on a Long Axial Field-of-View PET/CT Scanner", Diagnostics, vol. 13, no. 22, Art. no. 22, Jan. 2023, https://doi.org/10.3390/diagnostics13223418

ii H. Sari et al., "Feasibility of using abbreviated scan protocols with population-based input functions for accurate kinetic modeling of [18F]-FDG datasets from a long axial FOV PET scanner", Eur J Nucl Med Mol Imaging, vol. 50, no. 2, pp. 257–265, Jan. 2023, https://doi.org/10.1007/s00259-022-05983-7

[iii] D. Sarrut et al., ''The OpenGATE ecosystem for Monte Carlo simulation in medical physics'', Physics in Medicine and Biology, vol. 67, no. 18, p. 184001, Sept. 2022, https://doi.org/10.1088/1361-6560/ac8c83

[iv] C. M. Pommranz et al., ''A digital twin of the Biograph Vision Quadra long axial field of view PET/CT: Monte Carlo simulation and image reconstruction framework'', EJNMMI Phys, vol. 12, no. 1, p. 31, Mar. 2025, https://doi.org/10.1186/s40658-025-00738-3

[v] W. Lan et al., ''A Simulation Framework to Establish Ground Truth for Motion Correction in Total-Body PET: Initial Evaluation for Complex Respiratory Motion'', in 2024 IEEE Nuclear Science Symposium (NSS), Medical Imaging Conference (MIC) and Room Temperature Semiconductor Detector Conference (RTSD), Oct. 2024, pp. 1–2. https://doi.org/10.1109/NSS/MIC/RTSD57108.2024.10656116

[vi] W. P. Segars et al., ''4D XCAT phantom for multimodality imaging research'', Medical Physics, vol. 37, no. 9, pp. 4902–4915, Sept. 2010, https://doi.org/10.1118/1.3480985

[vii] P. J. Schleyer et al., ''Retrospective data-driven respiratory gating for PET/CT'', Physics in Medicine & Biology., vol. 54, no. 7, p. 1935, Mar. 2009, https://doi.org/10.1088/0031-9155/54/7/005

BWRSE4HPC - ENABLING SCIENTIFIC DISCOVERY THROUGH RSE SUPPORT

MARCEL KOCH[1]*, ANDREAS BAER[1], RENÉ CASPART[1], JASMIN HÖRTER[1], GLEN HUNTER[1], THOMAS ISENSEE[2], DOMINIC KEMPF[2], KAI RIEDMILLER[2], TIM SCHRADER[1] AND INGA ULUSOY[2]

[1]SCIENTIFIC COMPUTING CENTER, KARLSRUHE INSTITUTE OF TECHNOLOGY, KARLSRUHE, GERMANY

[2]SCIENTIFIC SOFTWARE CENTER, INTERDISCIPLINARY CENTER FOR SCIENTIFIC COMPUTING (IWR), HEIDELBERG UNIVERSITY, HEIDELBERG, GERMANY

* MARCEL.KOCH@KIT.EDU

ABSTRACT

Software has become a key enabler for modern science. The bwRSE4HPC initiative supports researchers in Baden-Württemberg who use software to achieve their research goals. The project embeds professional research software engineering (RSE) practices into scientific projects that utilize high-performance computing (HPC). We present two collaborative efforts where bwRSE4HPC helped improve the performance, maintainability, and sustainability of scientific software running on bwHPC infrastructure.

Keywords: Research software engineering, High-performance computing, Software development support, Software sustainability, Scientific collaboration

1. INTRODUCTION

Modern research increasingly depends on software, both commercial and custom-built. While many researchers use established tools such as computational fluid dynamics solvers, molecular dynamics simulators, or data analysis platforms, others rely on open-source software or develop their own in-house solutions. This often places a significant burden on researchers, who may lack formal training in software engineering and face pressure to prioritize scientific results over code quality. The frequent turnover of developers, typically PhD students or postdocs on fixed-term contracts, further exacerbates challenges in software maintainability and long-term usability. These issues are well documented in recent efforts to define the competencies and responsibilities of Research Software Engineers[i].

The newly established bwRSE4HPC[1] initiative addresses these issues by providing expertise in sustainable research software engineering. The bwRSE4HPC team contributes specialized skills in areas such as performance optimization, parallelization, build systems, and maintainability. Researchers based at higher education institutions in the German state of Baden-Würt-

1 https://www.bwrse4hpc.de/

temberg can apply for support in the form of short- to mid-term collaborative projects, typically lasting around six weeks or up to six months. The only requirement is that the software benefits from or is intended to run on high-performance computing (HPC) infrastructure.

In the following, two recent projects are described to highlight the benefits that are dedicated to RSEs can bring to research groups. In the first project, the RSEs improved the usability of the KIMMDY (**KInetic Monte Carlo/Molecular DYnamics**) software, and in the second project, the RSEs improve the sustainability and performance of the M++ software.

2. KIMMDY - REACTIVE MOLECULAR DYNAMICS ON GPUS

KIMMDY[ii],[2] is a kinetic Monte Carlo code that has been developed for molecular dynamics (MD). KIMMDY acts as an interface for different reactions and topologies to act on molecules. These processes can be provided through the plug-in interface of KIMMDY, allowing users to define their own plugins and making the code extensible. The molecular dynamics part of KIMMDY utilizes GROMACS, a highly parallelized molecular dynamics code for simulating the Newtonian dynamics of many particles. The software additionally makes use of GPU acceleration with CUDA to produce results efficiently.

The researchers using KIMMDY faced three issues: 1) Complicated dependency management, 2) Lacking confidence in plug-in integrations, and 3) Excessive use of GPU memory. During the collaboration, RSEs from bwRSE4HPC tackled these issues.

A. DEPENDENCY MANAGEMENT

The KIMMDY ecosystem consists of the main KIMMDY package and currently consists of five plugins. To simplify the dependency management, the project was restructured. The main repository links the officially supported plugins as sub-repositories. Furthermore, the recommended development setup now uses uv[3], and the plugins are registered in a uv-workspace. This workspace allows compiling the dependencies of all packages into one lock file, ensuring compatibility. The new workspace also speeds up installation compared to the previous setup using pip. Lastly, the switch to UV allows installing KIMMDY and all plugins, including the machine learning frameworks used, and the Python runtime, within a single command. This should make setup on the HPC infrastructure trivial. Support for the latest GROMACS version was added, further simplifying the process, as GROMACS no longer needs to be patched with PLUMED during compilation.

2 https://graeter-group.github.io/kimmdy/
3 https://docs.astral.sh/uv/

B. PLUG-IN CONFIDENCE

While KIMMDY itself has sufficient unit testing, the full simulation pipeline, including plug-ins, was less thoroughly tested. In particular, three out of the five plug-ins were missing integration tests. The RSEs implemented the integration tests for all missing plug-ins, following a similar framework for the integration tests that already exist within KIMMDY. The integration tests were tested on both Helix[4] and the CI pipeline is present with GitHub Actions.

One notable exception is that the integration tests of the KIMMDY-hydrolysis plug-in, which were only run on Helix. This is due to the integration tests taking too long for the CI pipeline. In addition to the development of the integration tests, testing frameworks for KIMMDY-hydrolysis and KIMMDY-dimerization were established for their respective CI pipelines. This had the additional benefit of increasing the total test coverage of KIMMDY from 4% to 77%.

C. GPU MEMORY

The plugin for the HAT reaction uses machine learning models to predict reactions. The researchers were experiencing an issue with GPU memory usage, in which the machine learning framework used, TensorFlow, was not releasing the GPU memory after the models were generated and used. When KIMMDY would start an MD simulation using GROMACS in the following step, GROMACS was unable to use the GPU due to insufficient memory available. As a consequence, twice as many GPUs were needed to run KIMMDY with the HAT plug-in..

TensorFlow itself and the implicit object management of Python don't provide a method to release GPU memory. The memory is only freed up completely when the parent process exits. Therefore, all code using TensorFlow was encapsulated in a separate process, which could be terminated after the predictions are done for a simulation. This incurs an overhead, since the models need to be reloaded frequently during a KIMMDY simulation. However, this was found to be negligible compared to the subsequent computations with GROMACS (~22 s vs. 12 h). As an additional layer of scrutiny, a unit test called test GPU memory release was implemented into the plug-in's testing framework.

3. M++ - INTEGRATION WITH LINEAR ALGEBRA LIBRARY

Solving linear systems is a critical part of the finite element (FE) library M++[iii]. Until now, these linear solvers were developed by the M++ team themselves, which led to an increased maintenance burden on the researchers, as well as less functionality than specialized libraries. Thus, the task of solving linear systems should be handed off to a specialized library.

4 https://wiki.bwhpc.de/e/Helix

The RSEs from bwRSE4HPC proposed to introduce the numerical linear algebra library Ginkgo[iv] as a backend for solving linear systems and generic linear algebra operations. The library provides GPU implementation for most of its components. Thus, it could enable more efficient use of M++ on modern supercomputers. Ginkgo is developed by a dedicated team, so M++ can benefit directly from their advances in numerical linear algebra. Only a small interface between M++ and Ginkgo is introduced, which delegates almost all operations to Ginkgo. Since the new backend will most likely not cover the full functionality of the existing backend, the Ginkgo backend will coexist with the existing backend.

As for writing this article, the project is still ongoing. After half of the project's runtime, the RSEs have succeeded in *1) Adding Ginkgo to M++'s build system, 2) Integrating Ginkgo solvers, and 3) Solving linear systems on a single GPU.*

The Ginkgo solver is part of the existing solver hierarchy in M++, which makes it easily available to users. Additionally, the external Ginkgo solver has a negligible overhead. Data from M++ is directly wrapped in Ginkgo data structures, which eliminates unnecessary copies at the interface. All single-node solvers and preconditioners from Ginkgo are available to M++. Ginkgo's multi-node components will be added in the second half of the project.

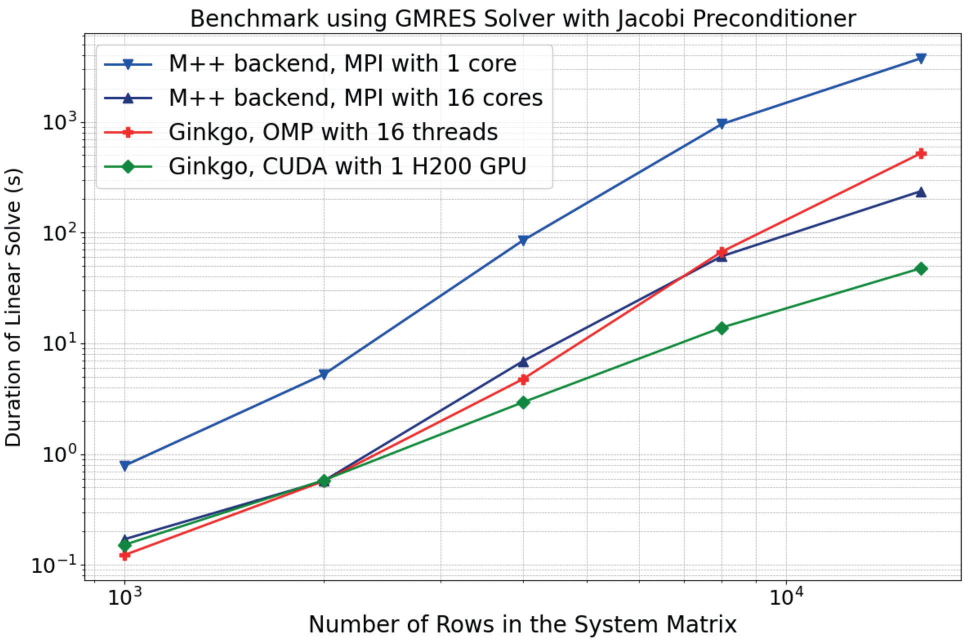


Figure 1: Time to solve a linear system comparing the built-in and Ginkgo backends. Run on a single HoreKa node

To highlight the benefit from the first phase of the project, a small benchmark example is considered. Fig. 1 shows the performance of the linear solver for configurations using M++ built-in backend and the Ginkgo backend. The linear system stems from a 2D Laplace problem, discretized by M++. All configurations use the GMRES solver with a Jacobi preconditioner. The benchmark is run on a single node of the HoreKa[5] supercomputer. While the CPU implementation of Ginkgo using OpenMP is comparable to M++'s implementation using MPI, the GPU implementation of Ginkgo provides the best time-to-solution for larger problems.

4. CONCLUSION

These two examples demonstrate how research in software engineering directly supports more efficient, scalable, and sustainable use of the bwHPC infrastructure. Tasks that are not directly tied to publishable outcomes can be addressed by dedicated RSEs. The KIMMDY project benefits from improved usability and better resource utilization, making it more attractive for other researchers. Although still ongoing, the M++ project already shows gains from leveraging modern GPU hardware while reducing the maintenance burden by relying on specialized libraries. Overall, bwRSE4HPC enables researchers without a strong computational background to successfully access and use bwHPC systems in their work. By applying best practices in software design, reproducibility, and automation, it fosters broader and more inclusive use of high-performance computing resources across Baden-Württemberg.

5. ACKNOWLEDGEMENT

bwRSE4HPC is funded by the Ministry of Science, Research, and Arts, Baden-Württemberg, Germany. The authors gratefully acknowledge the computing time provided on the high-performance computer HoreKa by the National High-Performance Computing Center at KIT (NHR@KIT). This center is jointly supported by the Federal Ministry of Education and Research and the Ministry of Science, Research and the Arts of Baden-Württemberg, as part of the National High-Performance Computing (NHR) joint funding program (https://www.nhr-verein.de/en/our-partners). HoreKa is partly funded by the German Research Foundation (DFG). The authors thank Niklas Baumgarten, Heidelberg University, for his input on M++, and Frauke Gräter, Jannik Buhr, and Eric Hartmann, Heidelberg University, for their input on KIMMDY.

5 https://www.nhr.kit.edu/userdocs/horeka/

6. REFERENCES

[i] F. Goth et al., ''Foundational Competencies and Responsibilities of a Research Software Engineer: Current State and Suggestions for Future Directions'', F1000Research, vol. 13, no. 1429, Sep. 2025, https://doi.org/10.12688/f1000research.157778.2

[ii] E. Hartmann, J. Buhr, K. Riedmiller, E. Ulanov, B. N. Schüpp, D. Kiesewetter, D. Sucerquia, C. Aponte-Santamaría and F. Gräter, ''KIMMDY: a biomolecular reaction emulator'', bioRxiv 2025.07.02.662624, July 2025, https://doi.org/10.1101/2025.07.02.662624

[iii] N. Baumgarten and C. Wieners, ''The parallel finite element system M++ with integrated multilevel preconditioning and multilevel Monte Carlo methods'', Computers & Mathematics with Applications, vol. 81, pp. 391-406, Apr. 2020, https://doi.org/10.1016/j.camwa.2020.03.004

[iv] Anzt et al., ''Ginkgo: A Modern Linear Operator Algebra Framework for High Performance Computing'', ACM Transactions on Mathematical Software, vol. 48, no. 1, pp. 1–33, Feb. 2022, https://doi.org/10.48550/arXiv.2507.11603

THOR: MASSIVELY PARALLEL GPU-ACCELERATED RADIATIVE TRANSFER

CHRIS BYROHL[1,2]* AND DYLAN NELSON[1,3]
[1]INSTITUT FÜR THEORETISCHE ASTROPHYSIK, HEIDELBERG UNIVERSITY, HEIDELBERG, GERMANY

[2]KAVLI IPMU, UNIVERSITY OF TOKYO, KASHIWA, JAPAN

[3]INTERDISZIPLINÄRES ZENTRUM FÜR WISSENSCHAFTLICHES RECHNEN, HEIDELBERG UNIVERSITY, HEIDELBERG, GERMANY

*CHRIS.BYROHL@UNI-HEIDELBERG.DE

ABSTRACT

Emission and absorption line features are important diagnostics for the physics underlying extragalactic astronomy. The interpretation of observed signatures involves comparing against forward-modeled spectra from galaxy formation simulations as well as more simplified geometries, including the complex scattering radiative transfer (RT) of resonant emission lines. We introduce THOR, a modern C++ Monte Carlo RT code focused initially on resonant emission lines. THOR is a high-performance, MPI-parallel distributed memory, SYCL-based multi-target code, running on CPUs, GPUs and other accelerators, yielding large 10-50x speed-ups compared to previous CPU-only codes. In this talk, we demonstrate THOR's science capabilities with several examples of use cases across scales, from simplified setups of the interstellar medium to cosmological simulations. Using THOR as a case study, we highlighted some of the lessons learned moving from an earlier code using openMP parallelism to a hybrid MPI+SYCL design for efficient use of accelerators.

Keywords: Monte Carlo methods, Radiative transfer, Astrophysics, SYCL, AdaptiveCpp

1. INTRODUCTION

Galaxies and their diffuse environments are observed through various emission and absorption features across the electromagnetic spectrum. Interpreting these observations lets us infer the distribution and physical state of matter across its phases and scales—from the interstellar to the intergalactic medium. Yet, this interpretation can be far from trivial, and we often rely on explicit forward-modeling of the radiative transfer (RT) on top of galaxy simulations.

RT modeling is particularly expensive for resonant and optically thick emission lines, where numerous scatterings drive up computational cost[i]. The computational expense could be significantly reduced by moving these calculations to GPUs. We therefore introduce **THOR**, a modern, GPU-accelerated C++ Monte Carlo RT code, initially focused on such resonant emission lines.

2. THEORY AND METHODS

We start from commonly available Monte Carlo-based RT implementations of resonant emission lines. Each Monte Carlo contribution represents a weighted photon package propagating through the spatial domain and sampling the respective probability distributions of physical interactions. This is done in a history-based approach, where each photon trajectory is computed independently on a given CPU thread via openMP parallelization. This strategy is readily ported to GPU threads, even though some adjustments are needed for optimization, as well as addressing hardware restrictions on the use of function pointers and dynamic memory allocation.

We heavily rely on C++ templating for flexible re-use of kernel code. We implement the RT code using the SYCL abstraction layer, enabling performance portability through modern C++ single-source code. SYCL is an open standard, and we opt for the Heidelberg-led AdaptiveCpp[ii] implementation to deploy the code across different accelerator vendors.

3. RESULTS AND DISCUSSION

The Lyα line of neutral hydrogen is not only one of the brightest emission lines in the Universe[iii], but one of the computationally most challenging resonant lines, thus making it a litmus test for THOR[iv]. Figure 1 shows a commonly used Lyα RT benchmark for different public

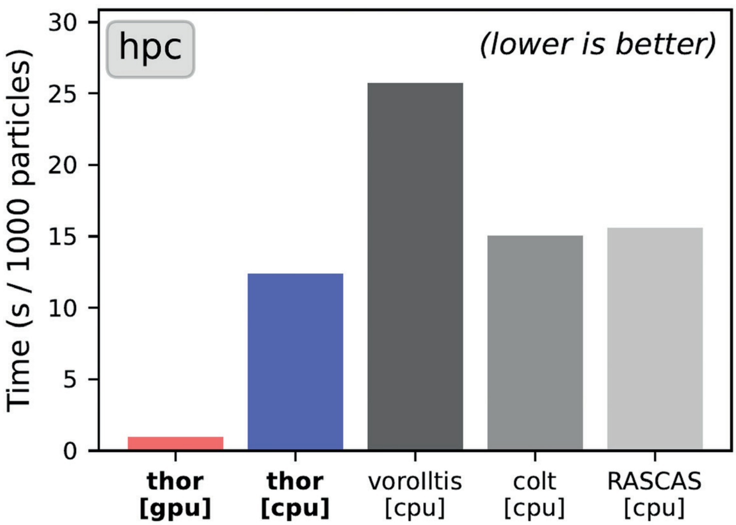


Figure 1: Code performance comparison of Thor on CPU and GPU backends against public codes. Here, we consider the Lyman-alpha Neufeld test problem in the optically thick limit. A single Nvidia A100 GPU was used compared to a 36-core Intel CPU.

resonant emission line RT codes, as well as THOR on a CPU and GPU backend. Overall, we find an order of magnitude speed-up of THOR on the GPU backend over the CPU backend.

We were able to successfully apply THOR on Intel, AMD, and Nvidia accelerators, showcasing the performance portability of THOR using the SYCL abstraction layer. We deployed the code on different HPC systems, including bwHPC binAC2. For all HPC systems, we needed to tweak the cluster environment and recompile the Clang compiler. This presents a significant hurdle for THOR users. We will explore the provision of user-space containers to mitigate this in the future, but also seek the help of system administrators to provide SYCL modules for these systems.

4. CONCLUSION

SYCL enables straightforward performance portability in modern C++. We successfully implemented a Monte Carlo radiative transfer code THOR, for galaxy simulation post-processing, initially focusing on resonant emission lines. We find speed-ups of order 10x or more for available GPUs over their CPU counterpart on HPC systems.

5. ACKNOWLEDGEMENT

This work was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy EXC 2181/1 - 390900948 (the Heidelberg STRUCTURES Excellence Cluster). CB and DN acknowledge funding from the Deutsche Forschungsgemeinschaft (DFG) through an Emmy Noether Research Group (grant number NE 2441/1-1). CB thanks the Marsilius Kolleg support for financial support for the required hardware. CB and DN also thank the Hector Fellow Academy for their funding support. Part of this work was carried out by CB as JSPS International Research Fellow. Calculations were carried out on the Vera, Raven and Viper(GPU) machines of the Max Planck Institute for Astronomy (MPIA) and systems at the Max Planck Computing and Data Facility (MPCDF). The authors acknowledge support by the High Performance and Cloud Computing Group at the Zentrum für Datenverarbeitung of the University of Tübingen, the state of Baden-Württemberg through bwHPC and the German Research Foundation (DFG) through grant no INST 37/1159-1 FUGG.

6. REFERENCES

i M. Dijkstra, "Physics of Lyα Radiative Transfer", in Saas-Fee Advanced Course, Springer-Verlag, M. Dijkstra, J. X. Prochaska, M. Ouchi, M. Hayes, A. Verhamme, P. North, S. Cantalupo, and H. Atek, Eds., in Saas-Fee Advanced Course, Berlin, Heidelberg: Springer, 2019, pp. 1–109. https://doi.org/10.1007/978-3-662-59623-4_1

ii A. Alpay and V. Heuveline, ''One Pass to Bind Them: The First Single-Pass SYCL Compiler with Unified Code Representation Across Backends'', in Proceedings of the 2023 International Workshop on OpenCL, in IWOCL '23. New York, NY, USA: Association for Computing Machinery, Apr. 2023, pp. 1–12. https://doi.org/10.1145/3585341.3585351

iii M. Ouchi, ''Observations of Lyα Emitters at High Redshift'', Saas-Fee Advanced Course, Springer-Verlag, vol. 46, p. 189, 2019, https://doi.org/10.1007/978-3-662-59623-4_3

iv C. Byrohl and D. Nelson, ''THOR: a GPU-accelerated and MPI-parallel radiative transfer code'', July 01, 2025, arXiv, https://doi.org/10.48550/arXiv.2507.11603

INTERACTIVE AI ON BWHPC – LESSONS LEARNED FROM BUILDING A LARGE-SCALE IMAGE ANALYSIS PLATFORM

ALEXANDER ZEILMANN[1,2]*, ERIK SCHNETTER[3] AND VINCENT HEUVELINE[1,2,3]

[1]INTERDISCIPLINARY CENTER FOR SCIENTIFIC COMPUTING, HEIDELBERG UNIVERSITY, HEIDELBERG, GERMANY

[2]DATA MINING AND UNCERTAINTY QUANTIFICATION GROUP, HEIDELBERG INSTITUTE FOR THEORETICAL STUDIES, HEIDELBERG, GERMANY

[3]UNIVERSITY COMPUTING CENTER, HEIDELBERG UNIVERSITY, HEIDELBERG, GERMANY
*ALEXANDER.ZEILMANN@IWR.UNI-HEIDELBERG.DE

ABSTRACT

AI and related technologies have rapidly reshaped the landscape of high-performance computing. Accordingly, many researchers in Baden-Württemberg face challenges due to the limited availability of interactive, GPU-ready, and service-oriented infrastructure within bwHPC. Drawing on systems we developed — KI-Morph and YoKI — we offer an experience-informed position on bwHPC's infrastructure remit, and the directions required to support modern research. While traditional, batch-oriented cluster infrastructure remains essential, we argue for a modular, interoperable approach that integrates clusters, cloud services, storage and more. We summarize today's landscape, introduce dimensions for diversification, and note that many of these capabilities already exist in mature industry platforms, underscoring feasibility and informing priorities within bwHPC. We note the risk that inaction will push researchers permanently to external providers and close with a call to collaborative enhancement of our infrastructure.

Keywords: AI, Interactive HPC, Interoperable Infrastructure, Cloud & Cluster Infrastructure Requirements

1. INTRODUCTION

Science has significantly changed in recent years in all areas from scientific exploration, interdisciplinary collaboration and outreach activities. In particular AI-driven research increasingly depends on interactive, GPU-accelerated, and service-oriented workflows and applications that cannot be supported by traditional batch systems. We experienced these challenges ourselves when developing KI-Morph[1], a platform for large-scale image analysis, and YoKI[2], the official LLM platform of Heidelberg University, as both applications require persistent avai-

1 https://ki-morph.de
2 https://www.urz.uni-heidelberg.de/de/service-katalog/kuenstliche-intelligenz-ki/yoki

lability, API access, workloads that come in bursts, and more. Accordingly, we want to share our perspective that draws on building and operating these systems as well as discussions with researchers and infrastructure teams. While the existing batch-oriented infrastructure remains a strength for many workloads and should be kept and developed further; nonetheless, new complementary capabilities are needed. We acknowledge that this is certainly not a simple task and will require significant work, time and funding. However, we believe that tackling these challenges is necessary and worth the effort, as this will make sure that we further support modern research in Baden-Württemberg and beyond.

The remainder of this paper surveys existing infrastructure, introduces key dimensions for infrastructure diversification, motivates modularity for interoperability, discusses the threat of inaction, and concludes with a call to collaborative modernization.

2. RETHINKING HPC IN THE AGE OF AI

Historically, almost all high-performance computing at our institutions was handled by Slurm-based batch computing. With new challenges and possibilities in the broader space of demanding computing, two major views of HPC have emerged. One view retains the historical meaning: Slurm-based, throughput-oriented batch computing on shared clusters, a perspective often held by technical experts. The other view broadens the term to encompass all highly demanding computing, including interactive, GPU-accelerated, and service-oriented workloads, a perspective often held by non-technical users. Neither perspective is incorrect; they reflect different vantage points. The historical view is solution-oriented, anchored in established tooling and operational models. The broader view is problem-oriented, centered on emerging user needs and application patterns. Depending on which view one adopts, opinions naturally diverge on the remit of bwHPC and whether the infrastructure we propose aligns with that remit. From our standpoint, the choice of definition and who operates the infrastructure matters less than ensuring that such infrastructure exists and is accessible. Because the proposed infrastructure addresses requirements that are pressing today, establishing it is more important than resolving terminological debates. Some may judge the proposal a misfit for bwHPC when using the classic definition. However, even if bwHPC is not the perfect organizational home, it is currently our closest initiative to the infrastructure we advocate, which motivates presenting our work in this context.

3. CURRENT BWHPC INFRASTRUCTURE LANDSCAPE

We first review the existing infrastructure landscape within bwHPC to contextualize the requirements that follow. We briefly assess cluster, cloud, and storage offerings and their suitability for interactive, AI-centric workloads.

While bwHPC clusters are familiar to most readers, the bwVisu remote visualization service built on top of them may be new. bwVisu enables interactive applications to run on cluster nodes by streaming their user interface to researchers. Because access remains Slurm-based, this approach does not suit persistent, service-oriented applications that must be always-on and API-accessible. However, because tools like bwVisu are well supported, clusters often appear more mature than cloud offerings and are used for projects that would be better served by cloud-native services. KI-Morph is one such example: we built on the Helix cluster via bwVisu, because it was possible and cloud alternatives were not yet ready, not because the cluster was the ideal infrastructure.

Today, three cloud approaches relevant to our community exist, but each is deficient for AI-centric, interactive workloads. bwCloud is intended as a shared cloud infrastructure for the bwHPC community, but it – to the best of our knowledge – currently does not contain GPU resources. Similarly, heiCLOUD, the cloud offering of Heidelberg University, does not yet provide GPU resources. The de.NBI Cloud[3] is a capable offering for the bioinformatics community in Germany, but its scope restriction prevents use for university-wide platforms such as YoKI or other AI services outside bioinformatics.

Across bwHPC, several storage solutions serve distinct purposes, including long-term archival, data exchange solutions, and storage of hot data for active computation. For hot data today, we effectively rely on a single class of solutions: the Large Scale Data Facility (LSDF) at KIT and the SDS Frontend at Heidelberg University. These solutions provide general-purpose storage and are well-suited for many research use cases. However, they can be ill-suited for very sensitive datasets, such as medical records, where specialized compliance and fine-grained access controls are required. They also do not easily support unauthenticated public file access, which some outreach projects require to distribute data openly.

4. DIMENSIONS FOR INFRASTRUCTURE DIVERSIFICATION

Rather than listing our own personal infrastructure requirements, we propose a broad problem-focused approach to enhancing the infrastructure. To this end, we believe that the requirements of users of the bwHPC infrastructure and, equally, the demands researchers outside of it should be gathered and organized. Concrete technical choices should follow from this analysis and be prioritized by urgency and complexity. Against that backdrop, we suggest considering demands and solutions to be diversely distributed along a variety of dimensions.

One important dimension is the amount of compute required. We believe that this is handled well by the established classification of HPC clusters into tiers, where bwHPC clusters

3 https://www.bw-cloud.org, https://heicloud.uni-heidelberg.de, https://cloud.denbi.de

(Tier 2) fit well within the broader landscape. Another dimension is alignment to scientific fields. While our clusters are organized by domain today, it is debatable whether this yields meaningful technical advantages. With research increasingly interdisciplinary, infrastructure organized strictly by scientific field may be a poor fit for many projects. A particularly relevant dimension for modern applications is the accessibility–security trade-off. Accessibility and security are not opposites, but raising one often constrains the other in practice. Some domains, especially medical research, require stringent controls, while outreach projects demand maximum ease of access. Further, applications also vary along the time-model dimension: runtime and start-up characteristics. Some workloads run for days and can tolerate long start-up times, while other workloads require persistent services and millisecond-scale bursts, for which we currently lack a suitable, integrated solution.

Two further dimensions concern openness and ownership. Open-source versus closed-source should remain an area of real variability; excluding closed-source options a priori can unnecessarily limit viable solutions. Similarly, owning and operating hardware on-premises versus renting resources that may or may not be located in-house should be considered without any preconceived bias. Both dimensions should follow from the technical and organizational requirements above, not precede them.

Additional dimensions include developer enablement and operational complexity, I/O throughput and latency for data-intensive workloads, and service-level objectives such as uptime availability guarantees. Cost and funding models, data locality and egress patterns, and compliance constraints can also be considered design dimensions. Choices along all these dimensions should be made in dialogue with researchers, operators, and policy makers to align trade-offs with real needs. For dimensions that have a significant variability in the demands, the infrastructure should reflect this variability holistically. We deliberately avoid prescribing a specific technical solution, instead, we emphasize that these dimensions should serve as a framework for building a future-facing diversification of our infrastructure.

5. DESIGN PRINCIPLES FOR MODULAR INFRASTRUCTURE

Besides these dimensions, we want to highlight the importance of modularity for the infrastructure. Modern research applications are composed of services that must interoperate across boundaries: data storage, compute, authentication, user-facing components and more. Creating isolated silos of infrastructure undermines reuse and raises operational burden in the long term. Instead, modularity should be a first-class design goal so that components can be composed for different projects and easily swapped as requirements evolve. Practically, this implies that clusters, cloud services and storage solutions are all able to be used together. The cloud can trigger jobs on the cluster, which can create files on the storage solution and that in turn can be accessed by the cloud. As the infrastructure evolves, and more components are added the modularity will become increasingly more important.

6. THE THREAT OF INACTION

It is tempting to adopt the narrow, traditional definition of HPC and continue investing solely in Slurm-based infrastructure. However, user requirements for a modernized infrastructure will not disappear by simply ignoring them. When researchers cannot satisfy these needs on bwHPC, they migrate to external providers that already offer suitable services. Once teams have established data pipelines, security reviews, and operational practices on external platforms, returning to bwHPC becomes costly and unlikely. Proactive investment in a more diverse, modular infrastructure is therefore essential to retain researchers and ensure long-term relevance.

7. CONCLUSION

The rise of AI and related technologies has rapidly expanded the community of researchers who need infrastructure tailored to interactive, data-intensive, and GPU-accelerated workflows. To meet these needs, we must diversify the infrastructure across clusters, cloud services, and storage, and make them modular and interoperable. Doing so will require sustained effort: structured interviews with researchers, careful evaluations, and open discussion between operators, policymakers, and domain experts. However, inaction is not an option; if needs are unmet, researchers will migrate to external platforms and may not return once their new workflows are established. We should embrace the changing HPC landscape as an opportunity to modernize, strengthen collaboration across institutions, and provide an infrastructure that enables cutting-edge, responsible research in Baden-Württemberg and beyond.

NUMERICAL SIMULATION OF THE AIRFLOW INSIDE DISC MOWERS TO PROTECT GRASSLAND INSECTS

JONAS FRANK[1]*, LEA VON BERG[2], STEFANIE ERHARDT[3], OLIVER BETZ[2], JOHANNES STEIDLE[3] AND STEFAN BÖTTINGER[1]

[1]DEPT. FUNDAMENTALS OF AGRICULTURAL ENGINEERING, UNIVERSITY OF HOHENHEIM, STUTTGART, GERMANY

[2]DEPT. EVOLUTIONARY BIOLOGY OF INVERTEBRATES, UNIVERSITY OF TÜBINGEN, TÜBINGEN, GERMANY

[3]DEPT. CHEMICAL ECOLOGY, UNIVERSITY OF HOHENHEIM, STUTTGART, GERMANY
*JONAS_FRANK@UNI-HOHENHEIM.DE

ABSTRACT

In recent years, the global decline in arthropod diversity and abundance has been linked to intensive agricultural practices, including large-scale monoculture and frequent mowing of grasslands. Disc mowers, widely used for their efficiency and low operating costs, are suspected to contribute to insect mortality due to the high rotational speed and slanted geometry of their blades, which may create a suction effect that pulls insects into the mower. To investigate air flow inside disc mowers, a simplified single-disc model with two blades and an underlying mower bar was developed and examined both experimentally, using a test rig at the Institute of Agricultural Engineering in Hohenheim, and numerically through CFD simulations in Ansys Fluent. Air velocity and pressure were recorded at five measurement points on a variable ground plate representing the meadow surface. Preliminary results indicate that the rotating blades generate highly turbulent airflow characterised by alternating low- and high-pressure zones, rather than a steady low-pressure field. This suggests that the air flow may not be the primary cause of insect loss. However, the simplified model excludes interactions between multiple discs and vegetation, which likely influence air flow. Future research will refine simulations and explore design modifications. As the influence of mowing techniques on insect loss seems smaller than previously assumed, unmown refuges gain importance as a habitat for insects that survived the mowing event.

Keywords: Numerical Simulation, Disc Mower, Insect Conservation, Grassland Management, Agricultural Engineering, High-performance Computing

1. INTRODUCTION

Over the past several years, a significant global decline in arthropod diversity has been observed[i]. Alongside factors such as climate change[ii], one of the major contributors to this trend is industrialised agriculture[iii]. Large-scale monoculture farming has led to widespread habitat loss and a reduction in structural diversity within ecosystems. In recent years, mowing has also been identified as an important factor for insect loss in agricultural grassland. Today, disc mowers are the prevalent technology for mowing agricultural grassland[iv,v], offering a combination of high performance, reliability and low operating costs. However, plots mown with a disc mower showed a significant reduction of individuals ranging from **28% for Hymenoptera to 55% for holometabolous larvae**, compared to an unmown control[vi]. This is often attributed to the high rotational speed of the blades (up to 80 km/h[vii]) and their slanted geometry, which in combination are proposed to generate a suction effect that pulls insects from the ground into the mower[viii, ix, x]. The evaluation and subsequent reduction of the suction effect is a main objective in the development of insect-friendly disc mowers.

Other technical solutions to improve disc mowers include insect flushing bars, which move through the grass in front of the mower and aim to chase away the insects. Effectiveness of insect flushing bars vary depending on the mobility of the insect groups present in the meadow[xi]. In contrast improving the suction effect could benefit all insects, independent of mobility.

2. THEORY AND METHODS

As direct measurements of the highly turbulent airflow inside rotating mowers are prone to measurement errors[xii], numerical simulations are a valuable tool for analysing air movement inside these machines. A simplified model, consisting of a single mower disc with two blades and the underlying mower bar, was developed and implemented both as a test rig at the Institute for Agricultural Engineering in Hohenheim and as a **CFD** simulation using Ansys Fluent[xiii]. A variable ground plate is located below the mower disc, representing the ground of the meadow. On this ground plate, five measurement points were specified, where air pressure and velocity were recorded in the test rig, as well as in the corresponding simulations.

Several modifications were implemented on the model, including higher cutting heights, a modified blade geometry and reduced rotating frequency. Due to the turbulent airflow and the high rotating frequency of the mower disc, the resulting transient simulation required a fine temporal resolution and large computational resources. Calculations were therefore conducted on the bwUniCluster 3.0 using the Ansys Fluent Software module.

3. RESULTS AND DISCUSSION

Preliminary results show that the rotation of the slanted blades in a standard setting of a disc mower leads to a highly turbulent air flow on the ground. This does not result in the formation of a permanent and stable low-pressure zone on the ground but creates a turbulent pattern with alternating low- and high-pressure zones. Low pressure is caused by the blades, whereas high pressure is caused by the flat side of the mower discs. The alternating low and high-pressure zones coincide with constantly changing air flow directions

This flow pattern suggests that the air flow created by the standard configuration of a disc mower cannot effectively lift insects from the ground into the mower. This finding is consistent with newer research, which shows that disc mowers cause similar reduction in insect abundance compared to other mowing techniques[xiv,xv,xvi].

However, the model used for the test rig and the simulation does only consist of one mower disc, where in reality there are multiple discs with overlapping rotation zones. Therefore, the airflow in real disc mowers is created by multiple mower discs and could potentially be even more turbulent. In addition, the simplified model does not contain any grass. The long, uncut grass in front of the mower, the stubble below the mower and the cut grass transported over the mower discs probably influence the airflow, which cannot be evaluated with the model implemented here.

4. CONCLUSION

Further work will aim to validate these results and explore their implications for mower design and insect conservation. As increased research shows that the influence of the mowing techniques on the loss of insects by mowing is smaller than previously assumed, the focus shifts to preserving unmown refuges. As most insect groups depend on the habitats created by a high grown meadow, the complete removal of the grass by mowing is a total destruction of their habitats. In addition to the insects in the refuges that were not affected by mowing at all, these refuges can serve as a habitat for the insects that survived the mowing. This is especially important when insect flushing bars are used, as the insects that escaped the mower need a habitat to live in after mowing.

5. ACKNOWLEDGEMENT

The project InsectMow is funded within the Federal Program Biodiversity (FKZ: 3520685A13) by the Federal Agency for Nature Conservation (BfN) with funds from the Federal Ministry for the Environment, Nature Conservation, Nuclear Safety and Consumer Protection (BMUV).

6. REFERENCES

i C. A. Hallmann et al., "More than 75 percent decline over 27 years in total flying insect biomass in protected areas", *PloS one*, early access. https://doi.org/10.1371/journal.pone.0185809

ii J. A. Harvey et al., "Scientists' warning on climate change and insects", *Ecological Monographs*, vol. 93, no. 1, 2023, Art. no. e1553, https://doi.org/10.1002/ecm.1553

iii F. Sánchez-Bayo and K. A. Wyckhuys, "Worldwide decline of the entomofauna: A review of its drivers", *Biological Conservation*, vol. 232, pp. 8–27, 2019, https://doi.org/10.1016/j.biocon.2019.01.020

[iv] J.-Y. Humbert, J. Ghazoul, N. Richner, and T. Walter, ''Hay harvesting causes high orthopteran mortality'', *Agriculture, Ecosystems & Environment*, vol. 139, no. 4, pp. 522–527, 2010. https://doi.org/10.1016/j.agee.2010.09.012

[v] D. van de Poel and A. Zehm, ,,Die Wirkung des Mähens auf die Fauna der Wiesen – Eine Literaturauswertung für den Naturschutz'', *Anliegen Natur*, vol. 36, no. 2, pp. 36–51, 2014.

[vi] L. von Berg, J. Frank, O. Betz, J. L. M. Steidle, S. Böttinger, and M. Sann, ''Disc mower versus bar mower: Evaluation of the direct effects of two common mowing techniques on the grassland arthropod fauna'', *Journal of Applied Ecology*, no. 62, 2025, Art. no. 1365-2664.14852, https://doi.org/10.1111/1365-2664.14852

[vii] L. von Berg, J. Frank, M. Sann, O. Betz, J. L. M. Steidle, and S. Böttinger, ,,Insekten- und spinnenschonende Mähtechnik im Grünland – Überblick und Evaluation'', (LANDTECHNIK, Bd. 78 Nr. 2 (2023)), 2023, https://doi.org/10.15150/lt.2023.3291

[viii] D. Ehlert, ,,Schlegelmulchgeräte'', *Landtechnik*, vol. 48, no. 12, pp. 643–645, 1993.

[ix] M. Lösch, D. Strauß, H. Wandel, and T. Jungbluth, ,,Der Bio-Cutter: Ein technischer Ansatz zur faunaschonenden Mulchtechnik'', *Landtechnik*, vol. 52, no. 1, pp. 18–19, 1997.

[x] S. Rux, ,,Geräte und Verfahren zur Brachflächen- und Landschaftspflege: Mäh- und Mulchgeräte'', in *Landtechnik: 136 Tabellen* (Landwirtschaftliches Lehrbuch), H. Eichhorn, Ed., 7th ed. Stuttgart: Ulmer, 1999, pp. 426–428.

[xi] L. von Berg et al., ''Minimising insect mortality during grassland mowing: The potential of insect chasing devices'' ,*Insect Conservation and Diversity*, 2025, Art. no. icad.12854, https://doi.org/10.1111/icad.12854

[xii] D. Ehlert, ,,Schlegelmulchgeräte'', *Landtechnik*, vol. 48, no. 12, pp. 643–645, 1993.

[xiii] *Ansys Fluent 2024 R2* (2025). Canonsburg, PA, USA: Ansys Inc.

[xiv] L. von Berg, J. Frank, O. Betz, J. L. M. Steidle, S. Böttinger, and M. Sann, ''Disc mower versus bar mower: Evaluation of the direct effects of two common mowing techniques on the grassland arthropod fauna'', *Journal of Applied Ecology*, no. 62, 2025, Art. no. 1365-2664.14852, https://doi.org/10.1111/1365-2664.14852

[xv] D. Kraut, ,,Schädigung der Kleintierfauna durch Mähwerke'', *Landtechnik*, vol. 50, no. 3, pp. 138–139, 1995.

[xvi] C. Schwarz, F. Fumy, M. Drung, and T. Fartmann, ''Insect-friendly harvest in hay meadows – Uncut refuges are of vital importance for conservation management'', *Global Ecology and Conservation*, vol. 48, e02731, 2023, https://doi.org/10.1016/j.gecco.2023.e02731

INFLUENCE OF TRIAXIALITY ON THE DYNAMICS OF TRIPLE SUPERMASSIVE BLACK HOLES IN A COSMOLOGICAL CONTEXT

NAVONIL SAHA[1]*, PETER BERCZIK[2,3], ANDREAS JUST[1] AND MARGARITA SOBOLENKO[2,3]

[1]ZENTRUM FUR ASTRONOMIE DER UNIVERSITAT HEIDELBERG, ASTRONOMISCHES RECHEN-INSTITUT, HEIDELBERG, GERMANY

[2]NICOLAUS COPERNICUS ASTRONOMICAL CENTRE, POLISH ACADEMY OF SCIENCES, WARSAW, POLAND

[3]MAIN ASTRONOMICAL OBSERVATORY, NATIONAL ACADEMY OF SCIENCES OF UKRAINE, KYIV, UKRAINE

*NAVONIL.SAHA@UNI-HEIDELBERG.DE

ABSTRACT

The hierarchical nature of galaxy formation in the ΛCDM framework often leads to multiple supermassive black holes (SMBHs) in the galactic nuclei. The timescale over which galaxies merge, plays a crucial role in shaping the dynamical evolution and the merger dynamics of their central SMBHs. While binary SMBH evolution is well studied, the long-term dynamics of triple SMBH systems, particularly in non-spherical potentials, remain less understood. We investigate the role of triaxiality in the evolution and dynamics of triple SMBHs with initial conditions drawn from the ROMULUS25 cosmological simulation, using high-resolution gravitodynamical N-body simulations. We explore different orbital configurations and host shapes, tracking the evolution from galactic inspiral to hard binary formation at sub-parsec scales. In all cases, the two most massive SMBHs form a rapidly hardening binary that coalesces within a fraction of a Hubble time, while the third forms a stable hierarchical triple system with the heavier binary or remains on a wide orbit.

Keywords: Black hole physics, Supermassive black holes, Galaxies' kinematics and dynamics

1. INTRODUCTION

During their lifetimes, most massive galaxies undergo various major merger events with other galaxies. This has been the basis of the hierarchical cosmic structure formation in the ΛCDM cosmology[i,ii] Most elliptical and spherical galaxies host a central supermassive black hole (hereafter, SMBH), and they gain mass through hierarchical mergers. The emperical scaling relations indicate that the host galaxies and their central SMBHs grow together and suggest a co-evolutionary link between them, like the M_{BH}–σ relation, which connects the SMBH mass

(M_{BH}) to the stellar velocity dispersion (σ) of the galactic bulge[iii]. When the galaxy grows, σ increases and hence subsequently the M_{BH}.

The stellar dynamical evolution of the SMBH is commonly divided into three distinct stages, governed by different physical mechanisms[iv]. First, in the post-merger phase of galaxy mergers, the SMBHs sink to the center of the galaxy through dynamical friction with other stellar bodies and particles[v]. The SMBHs get decelerated by gravitational drag, causing them to lose orbital energy and angular momentum. This process operates efficiently on kiloparsec to parsec scales and continues until the two SMBHs form a bound binary. Once the SMBHs form a bound binary, further orbital decay is dominated by interactions with the surrounding stellar environment, and this is known as the stellar hardening phase. The pair thus forms a hard binary and continues ejecting stellar particles through three-body encounters (as a gravitational slingshot), carrying away energy and angular momentum and causes the binary to shrink[vi]. Finally, at sub-parsec scales, the emission of gravitational waves dominates, rapidly driving the binary to coalescence[vii].

The timescale from the merger of galaxies to the final coalescence of their central SMBHs can vary significantly, depending on the remnant's matter distribution and the binary's orbital parameters, ranging from tens of Myr to nearly a Gyr or longer[viii]. During this evolution, it is possible that a third galaxy joins this galaxy, which already contains an SMBH binary and creates a triple SMBH system. However, the long-term dynamics of triple SMBH systems still remain poorly understood. In previous work by[ix] triple SMBH systems, where two subsequent mergers take place within a Gyr, were extracted from the merger trees of SMBHs, and it was found that the triple interactions play a crucial role in the outcome of the dynamical evolution of the SMBHs. In these simulations spherical galaxy systems were used with the initial conditions taken from the ROMULUS25 cosmological simulation[x] and it turned out that the two most massive SMBHs merge first. In this work, we extend these models on triple SMBHs by investigating the role of triaxiality in shaping the dynamical evolution of three SMBHs systems to find the common dynamical evolution pattern of the triples and predict the typical coalescence time scale using high-resolution gravitodynamical N-body simulations.

2. METHODS AND INITIAL CONDITIONS

We use the spherical galaxy models of[ix] in order to create different models of 3-galaxy triaxial initial conditions with dark matter, gas and stars, taking the 3 galaxy radial galactic Dehnen profile parameters along with the SMBH initial conditions from Table 2 and Table 3 of the paper. The masses of the SMBHs are $88.4 \mathrm{x} 10^7 M_\odot$, $13.3 \mathrm{x} 10^7 M_\odot$, and $3.6 \mathrm{x} 10^7 M_\odot$, respectively, for BH1, BH2 and BH3. All three models A, B and C, in this work (corresponding to model A1, A2 and A3 in[ix]) differ in the initial galactic density profiles of the stellar components. We then add the triaxial shape taken from[xi] to these galaxies using AGAMA galaxy modeling software[xii]. Here, the potential of the SMBHs was not taken into account in order to create, within a few dynamical time scales of the nucleus in the simulation, a central cusp of the central potential of the SMBH.

We use different high-resolution N-body codes for our simulations. For the initial 3-body interaction phase, we start with the BONSAI2 Tree code[xiii] with ~10 million particles for each galaxy for sufficient interaction of the particles during the early evolution of the global galaxy merger phase and the decay of the SMBHs to the center of the merger remnant due to dynamical friction. Once the binary starts to harden and the separation of one black hole pair falls below ~100 pc, the risk of close pericenter passages not being properly resolved due to the lack of uniform timestep starts to grow, and we get a time resolution problem. Therefore, at this point, we shift to the φ-GPU Direct N-body code[xiv] with an individual black timestep scheme for the hardening phase of SMBHs and correctly resolve the 3-body dynamics close to SMBHs until we get a bound system. In this step, we also reduce the total number of particles from $30x10^6$ to $4.8x10^6$ particles close to the inner 10 kpc of the SMBHs. We also include Post-Newtonian (PN) approximation up to 2.5 PN order when the SMBHs reach a separation of $1000R_s$ (Schwarzschild radius) and move towards the relativistic regime.

3. RESULTS AND CONCLUSION

The orbital evolution of the triple SMBH system showed that they form a binary that orbits around the most massive BHs. From time to time, there are three-body encounters of the SMBHs, which switch the members of the binary for all our models (see Figure 1).

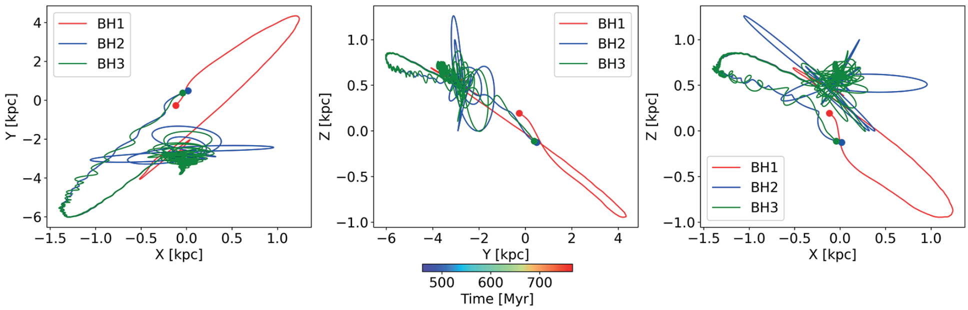


Figure 1: Orbital evolution of the three SMBHs, where the three BH1, BH2 & BH3 are color-coded as red, blue and green, respectively, showing the continuous exchange of the SMBHs due to three-body encounters

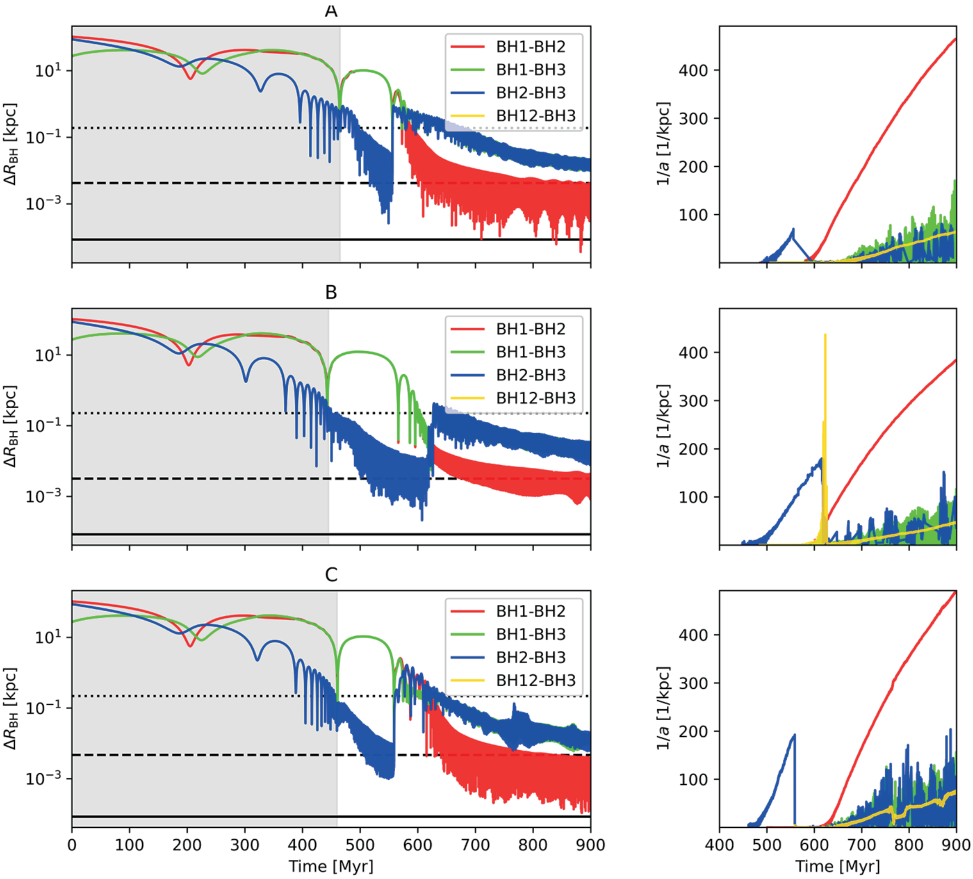


Figure 2: Black hole orbital parameters across different simulations. The color coding refers to different SMBH pairs. The red lines refer to the orbital parameters around the inner BH1-BH2 binary (the two most massive BHs). Background colors mark the codes used during the particular simulation period, light grey for Bonsai2, white for ϕ-GPU. The left panels show the separation of different SMBH pairs, and the right panels show the evolution of the inverse semi-major axis. The black dotted, dashed, and solid lines mark the influence radius, hardening radius, and the factor 1000 of the Schwarzschild radius (R_s) of BH1, respectively.

For all three models, initially, the BH2 and BH3 approach each other very closely, which then, later during the final galactic collision, weakens the BH2 slingshot. Hence, the BH2 immediately returns to the galactic center, and we get a BH2-BH3 system (blue line) sharing a common halo and orbiting each other, preventing the formation of the BH1–BH3 system (green line). Now, as the separation starts to fall below sub-pc scale, we switched to the direct code between 400-500 Myr. Between 500-700 Myr, there is a collision of the BH2-BH3 with the first galaxy, and we get an exchange and finally a BH1–BH2 hard binary (red line) forms, as BH3 is sent off to orbit around the galactic center at 100–500 pc (see Figure 2). While the

BH1–BH2 binary continues to harden, BH3 slowly descends, and it becomes bound to the inner BH1–BH2 hard binary and a stable hierarchical triple system is formed in all the models.

We take the orbital-averaged hardening rate by GW emission from the empirical formula given in[vii] assuming a constant eccentricity and add the stellar hardening estimated from the simulation, and calculate the merging time of the binary. We get a stellar hardening rate between 1-2 $kpc^{-1}Myr^{-1}$ for the hard binary and merging time between 3-13 Gyr, which is well within the Hubble time. The real merging rate of the models is different because of the different hardening rate of the binary system, which actually depends on the details of the orbit. Hence, we conclude that triaxiality does not show any strong impact on the interaction and the final outcome of the triple system as compared to SMBHs present in spherical galaxies.

4. ACKNOWLEDGEMENT

The authors gratefully acknowledge the Gauss Center for Supercomputing eV[1] for funding this project by providing computing time on the GCS Supercomputer JUWELS at Jülich Supercomputing Center (JSC). The authors acknowledge support by the High Performance and Cloud Computing Group at the Zentrum für Datenverarbeitung of the University of Tübingen, the state of Baden-Württemberg through bwHPC and the German Research Foundation (DFG) through grant no. INST 37/1159-1 FUGG, and the data storage service SDS@hd supported by the Ministry of Science, Research and the Arts Baden-Württemberg (MWK) through grant no. INST 35/1503-1 FUGG. NS acknowledges the support of doctoral scholarship from the German Academic Exchange Service (DAAD) through grant no. 91864951. PB and MS thank the support from the special program of the Polish Academy of Sciences and the U.S. National Academy of Sciences under the Long-term program to support Ukrainian research teams, grant No. PAN.BFB.S.BWZ.329.022.2023.

5. REFERENCES

i Peebles, P. J. E. and Yu, J. T., "Primeval Adiabatic Perturbation in an Expanding Universe", The Astrophysical Journal, vol. 162, IOP, p. 815, 1970. https://articles.adsabs.harvard.edu/pdf/1970ApJ...162..815P

ii White, S. D. M. and Frenk, C. S., "Galaxy Formation through Hierarchical Clustering", The Astrophysical Journal, vol. 379, IOP, p. 52, 1991. https://articles.adsabs.harvard.edu/pdf/1991ApJ...379...52W

iii Ferrarese, L. and Merritt, D., "A Fundamental Relation between Supermassive Black Holes and Their Host Galaxies", The Astrophysical Journal, vol. 539, no. 1, IOP, pp. L9–L12, 2000. https://iopscience.iop.org/article/10.1086/312838/pdf

iv Begelman, M. C., Blandford, R. D., and Rees, M. J., "Massive black hole binaries in active galactic nuclei", Nature, vol. 287, no. 5780, pp. 307–309, 1980. https://www.nature.com/articles/287307a0.pdf

v Chandrasekhar, S., "Dynamical Friction. I. General Considerations: the Coefficient of Dynamical Friction.", The Astrophysical Journal, vol. 97, IOP, p. 255, 1943. https://articles.adsabs.harvard.edu/pdf/1943ApJ....97..255C

vi Quinlan, G. D., "The dynamical evolution of massive black hole binaries I. Hardening in a fixed stellar background", New Astronomy, vol. 1, no. 1, Elsevier, pp. 35–56, 1996. https://doi.org/10.1016/S1384-1076(96)00003-6

1 https://www.gauss-centre.eu/

vii Peters, P. C. and Mathews, J., ''Gravitational Radiation from Point Masses in a Keplerian Orbit'', Physical Review, vol. 131, no. 1, APS, pp. 435–440, 1963. doi:10.1103/PhysRev.. 131.435. https://journals.aps.org/pr/pdf/10.1103/PhysRev.131.435

viii Khan, F. M. et al., ''Mergers of Unequal-mass Galaxies: Supermassive Black Hole Binary Evolution and Structure of Merger Remnants'', The Astrophysical Journal, vol.749, no. 2, Art. no.147, IOP, 2012.https://iopscience.iop.org/article/10.1088/0004-637X/749/2/147

ix Koehn, H., Just, A., Berczik, P., and Tremmel, M., ''Dynamics of supermassive black hole triples in the ROMULUS25 cosmological simulation'', Astronomy and Astrophysics, vol. 678, Art. no. A11, EDP, 2023. https://doi.org/10.1051/0004-6361/202347093

x Tremmel, M., ''The Romulus cosmological simulations: a physical approach to the formation, dynamics and accretion models of SMBHs'', Monthly Notices of the Royal Astronomical Society, vol. 470, no. 1, OUP, pp. 1121–1139, 2017. https://doi.org/10.1093/mnras/stx1160

xi Butsky, I., ''NIHAO project II: halo shape, phase-space density and velocity distribution of dark matter in galaxy formation simulations'', Monthly Notices of the Royal Astronomical Society, vol. 462, no. 1, OUP, pp. 663–680, 2016. https://doi.org/10.1093/mnras/stw1688

xii Vasiliev, E., ''AGAMA: action-based galaxy modelling architecture'', Monthly Notices of the Royal Astronomical Society, vol. 482, no. 2, OUP, pp. 1525–1544, 2019. https://doi.org/10.1093/mnras/sty2672

xiii Bédorf, J. et al., ''A sparse octree gravitational N-body code that runs entirely on the GPU processor'', *Journal of Computational Physics*, vol. 231, no. 7, Elsevier, pp. 2825–2839, 2012. https://doi.org/10.1016/j.jcp.2011.12.024

xiv Berczik, P., ''High performance massively parallel direct N-body simulations on large GPU clusters''., in *International conference on High Performance Performance Computing*, 2011, pp. 8-18. https://ui.adsabs.harvard.edu/abs/2011hpc..conf....8B/abstract

PALEOPLASIM 1.0: AN EARTH SYSTEM MODEL OF INTERMEDIATE COMPLEXITY FOR PALEOCLIMATE MODELING AND LARGE ENSEMBLE STUDIES

MURIEL RACKY[1]*, ARYA SAMANTA[1] AND KIRA REHFELD[1]

[1]GEO- AND ENVIRONMENTAL CENTER (GUZ), UNIVERSITY OF TÜBINGEN, TÜBINGEN, GERMANY

*MURIEL.RACKY@UNI-TUEBINGEN.DE

ABSTRACT

Here, we describe the first version of PaleoPlaSim, a paleoclimate-enhanced version of the Planet Simulator. PaleoPlaSim is a simplistic but fully coupled Atmosphere-Ocean General Circulation Model. Two dynamic vegetation models and a sea ice model are available as optional components. Our development work includes improved and additional parametrizations, adjustments to enable transiently changing boundary conditions, and accompanying scripts and documentation.

PaleoPlaSim can realize about 400 simulation years per day on a single 2.8 GHz AMD EPYC™ 7543 CPU core at a resolution of ~500km in the atmosphere. Shorter simulation times can be achieved through parallelization. This makes it ideally suited for long (paleoclimate) simulations, large ensembles, and idealized simulations that address fundamental research questions.

Here, we used PaleoPlaSim for a 250-member perturbed parameter ensemble, varying 12 model parameters. We analyzed the present-day and past climatologies and their relationship to climate variability. We identify a decoupling of local mean temperature change and variability in the tropics. This points to differing state-and-latitude-dependent climate feedback. Finally, we give an outlook on a Bayesian calibration procedure that considers multiple climate states and variability targets.

Keywords: EMIC, Paleoclimate modeling, Model calibration, Perturbed parameter ensemble

1. INTRODUCTION

Climate modeling utilizes a hierarchy of models from conceptual to complex. Earth System Models of Intermediate Complexity (EMICs) sit in the gap between simple energy balance models and Atmosphere/Ocean General Circulation Models (GCMs). With high computational efficiency, they are ideal for large ensembles and long-term simulations[i]. EMICs have low spatial resolution and use simplified expressions to model unresolved, sub-grid-scale processes. Such parametrizations of uncertain parameters are estimated through model calibration (denoted) tuning based on selected observations of climate variables[ii]. A perturbed parameter ensemble (PPE) enables the evaluation and optimization of these parameters by varying the uncertain parameters to realize a multitude of model variants. In the following, we present the EMIC PaleoPlaSim, which is based on PlaSim[iii] and a 250-member PPE. We outline future work for an iterative Bayesian tuning procedure.

2. MODEL OVERVIEW

PaleoPlaSim combines the Planet Simulator[iii] (PlaSim) with the Large-Scale Geostrophic Ocean General Circulation Model[v] (LSG). PlaSim and LSG exchange momentum, heat, and fresh water fluxes via a shared uppermost ocean layer[v]. The dynamical core of PlaSim is the Portable University Model of the Atmosphere[vi] (PUMA). PUMA is a simplified atmospheric GCM that solves the primitive equations for momentum, mass, and energy at spectral resolutions of T21 (~500km) and higher with ten or more vertical layers. LSG also solves the primitive equations but neglects the nonlinear terms of the Navier–Stokes equation based on a scale analysis[vii]. This filters out Rossby and gravity waves, which are believed to be less important on climate timescales. LSG operates on a 2.5°×5° semi-staggered Arakawa E grid with 22 vertical levels. Ocean diffusivity in LSG, radiative transfer, and phenomena of the hydrological cycle, soil physics, and dry convection in PlaSim[vi] are parameterized. Further sub-modules include a thermodynamic zero-layer sea ice model[ix] and two dynamic vegetation models (SEDGES[x] and SimBa[viii]). PaleoPlasim is at the top of the EMIC hierarchy with a dynamical atmosphere and ocean model, albeit with simple parametrizations and low spatial resolution.

3. MODEL DEVELOPMENT

PaleoPlaSim is based on the openly accessible PlaSim (v0318). The model is largely under the GNU General Public License 2.0, with source code accessible from the original developers at the University of Hamburg website[1]. PlaSim was a tool developed for theoretical meteorology and teaching, limited to equilibrium solutions. PaleoPlaSim now allows for transient time-varying boundary conditions in surface fields (e.g., land-sea mask, topography, and glacier mask), greenhouse gas concentration, and orbital configuration. Parametrizations have been improved or added: Albedo parameterizations for snow-covered surfaces now consider snow aging and surface properties. A latent heat correction for sea ice was introduced. Changes in aerosol optical depth (e.g., to simulate impacts of volcanic eruptions) and additional freshwater flow into oceans (“hosing”) can now be prescribed. Vertical diffusion in the ocean was changed. Spatially resolved friction coefficients can now be read from a file. These dimen-

1 https://www.mi.uni-hamburg.de/en/arbeitsgruppen/theoretische-meteorologie/modelle/plasim.html

sionless coefficients are multipliers in the prognostic equations for tracers and emulate physical barriers. We updated and added runscripts for configuring, running, and post-processing simulations, along with documentation. A full list of changes and options (e.g., clouds/deep convection; sea-ice, water isotopologues) is published with the code repository[2].

4. COMPUTATIONAL REQUIREMENTS AND BENCHMARKING

The following software is necessary to run the model: A FORTRAN-90 compiler, a make utility, an openMPI environment with compilers for parallel execution, the NetCDF library, CDO (Climate Data Operators), and a bash shell. The optional graphical user interface requires C and C++ compilers. PaleoPlaSim can realize about 400 simulation years per day on a single 2.8 GHz AMD EPYC™ 7543 CPU. In parallel execution, data arrays are partitioned along latitudes. The number of latitudes must be divisible by the number of processors with no remainder. For instance, in a T21 simulation with 32 latitudes, one could parallelize using 1, 2, 4, 8, 16, or 32 CPUs. Simulation years per day increase accordingly (Figure 1). For comparison, a GCM with 160/40 km atmospheric/oceanic resolution simulates ~85 years per day on 40 nodes, each with 128 cores, on the HLRE-4 supercomputer[xi]. PaleoPlaSim simulations are reproducible because the restart files created after each simulation year are hardware independent. Tests across machines show that simulated climatology depends solely on the experimental settings.

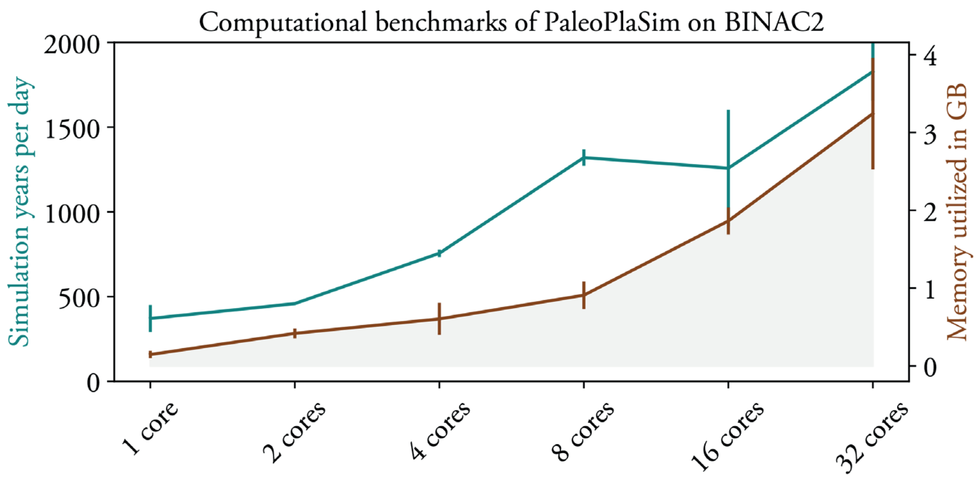


Figure 1: Number of simulation years realized per day (blue, left Y-Axis) and memory usage (brown, right Y-Axis) versus number of cores, tested on the BINAC2 compute partition

2 Published on zenodo: http://doi.org/10.5281/zenodo.17350864

5. RESULTS FROM PERTURBED PARAMETER ENSEMBLE

We created a 250-member PPE of PaleoPlaSim varying 12 model parameter (Table 1). Selected parameters were identified as key parameters drawing on feature importance sampling[xii,xiii]. For each PPE configuration, we simulate present-day (PD) and Last Glacial Maximum (LGM) equilibria for 1250 years. The global mean temperature difference between the two states (LGM-PD) is an indicator of the climate sensitivity[xiv,xv] of the respective PPE member.

Parameter	Related climate process	Description	Lower bound	Upper bound
gamma	Rain/convection (PlaSim)	Evaporation of precipitation in the atmosphere	0	1
tswr1	Radiation (PlaSim)	Cloud albedo	0,01	0,2
tswr2	Radiation (PlaSim)	Cloud backscattering	0,01	0,5
tswr3	Radiation (PlaSim)	Cloud single scattering albedo	0,001	0,01
acllwr	Radiation (PlaSim)	Mass absorption coefficient of clouds	0	0,2
drhssea	Ocean surface (PlaSim)	Evaporation properties of ice free ocean	0	10
albsmin	Land processes (PlaSim)	Snow albedo minimum value	0,4	0,65
albsmax	Land processes (PlaSim)	Snow albedo maximum value	0,8	0,9
arange	Vertical diffusion (LSG)	Range of vertical diffusion profile	0	3,00E-0,4
astar	Vertical diffusion (LSG)	Position of vertical diffusion profile	arange	1,00E-03
zstar	Vertical diffusion (LSG)	Position of vertical diffusion profile	0	6000
lambda	Vertical diffusion (LSG)	Shape of vertical diffusion profile	1,00E-06	1

Table 1: Description of the chosen 12 model parameters that were randomly varied within their bounds to create the PPE

The relation between the LGM-PD temperature anomaly and near-surface air temperature variability (Figure 2) shows that, globally, colder ensemble members exhibit a larger LGM-PD warming and higher temperature variability (Figure 2a). This indicates a coupling between the mean state and its variability, consistent with data-based results[xvi] and theoretical expectation from linearized energy balance models[xvii,xviii]. We hypothesize that in our ensemble, this is primarily due to varying positive climate feedback, such as the ice-albedo effect. However, regionally, particularly in the tropics, this relationship is weak and may be of opposing sign (Figure 2b). LGM to PD warming has been reconstructed based on data from environmental archives to be within 3 and 8K[xix,xx]. 72% of the ensemble configurations show warming within this range. Yet their within-state tropical temperature variability varies from 0.14 °C (standard deviation over the last 250 years of simulation) to 0.27 °C. This indicates that regional temperature variability is not fully determined by climate sensitivity, and that additional degrees of freedom could be explored for model calibration by comparison with paleoclimate data.

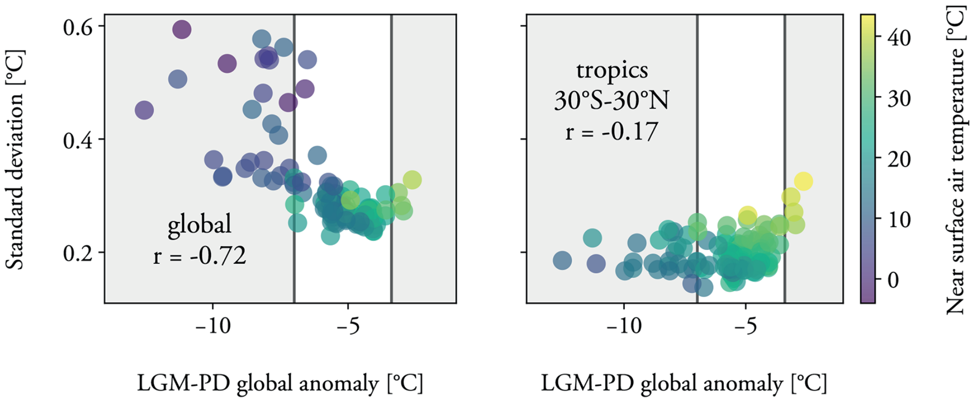


Figure 2: Global mean and near-surface air temperature variability (Y-axis) versus the LGM-PD temperature anomaly (X-axis) for the PPE. Each point corresponds to a PPE member, and its color indicates the mean temperature. Panel a) refers to the global temperature mean and standard deviation, while panel b) refers to the tropical latitudes (30°S-30°N). Pearson correlation coefficients r is given on the panels. The shaded area on the X-axis brackets the realistic range for LGM-PD warming.

6. OUTLOOK

We presented here a first proof-of-concept for an efficient model of Earth's climate over long timescales, PaleoPlaSim. In future, we plan to use our PPE simulations as the basis for an iterative Bayesian tuning procedure[xxi] combining present-day observations, LGM climate reconstructions, and a variability term in the tuning target. This exploratory approach should identify model parameters which are state-dependent or linked to climate variability.

The true Earth system is far more complex than our model, and a better process representation could improve it. For example, a dynamic sea-ice component was developed[xxii], and is so far not included. Nevertheless, we expect this version of PaleoPlaSim to be already useful for fundamental research on Earth's climate, climate modeling, and education. With its simplified yet complex character, PaleoPlaSim fills a niche between EMICs that fully neglect atmosphere and/or ocean dynamics and fully coupled GCMs. PaleoPlaSim could, for example, be used to develop fully automated model calibration, which could later be applied for more complex climate models.

7. ACKNOWLEDGEMENTS AND AUTHOR CONTRIBUTIONS

We thank everyone who was also involved in the development, testing, and documentation of PaleoPlasim: Heather Andres (core developer), Moritz Adam (development and advice), Olga Erokhina (documentation), Oliver Mehling (surrounding scripts), Pablo Paiewonsky (development of SEDGES), Jost von Hardenberg (postprocessing scripts), Audrey Parnell (testing), Felix Pollak (development), Leoni Rath (development), Irene Trombini (testing), and Frank Lunkeit (advice). We thank Heather Andres, Elisa Ziegler and Olga Erokhina for their valuable feedback on the manuscript. This work was conducted on bwUniCluster 2.0, 3.0, and BINAC2 and we gratefully acknowledge the support of the State of Baden-Württemberg through the bwHPC project. We acknowledge funding through the German Research Foundation (DFG, project no. 395588486), the federal ministry for education and research (BMBF, now BMFTR; project no 01LS2103A) and the University of Tübingen excellence funds.

8. REFERENCES

i S. L. Weber, "The utility of Earth system Models of Intermediate Complexity (EMICs)", WIREs Climate Change, vol. 1, no. 2, pp. 243–252, 2010, https://doi.org/10.1002/wcc.24

ii F. Hourdin et al., "The Art and Science of Climate Model Tuning", Bulletin of the American Meteorological Society, vol. 98, no. 3, pp. 589–602, 2017, https://doi.org/10.1175/BAMS-D-15-00135.1

iii K. Fraedrich et al., "The Planet Simulator: Towards a user friendly model", metz, vol. 14, no. 3, pp. 299–304, 2005. https://www.schweizerbart.de/papers/metz/detail/14/54202

iv E. Maier-Reimer et al., "Mean Circulation of the Hamburg LSG OGCM and Its Sensitivity to the Thermohaline Surface Forcing", Journal of Physical Oceanography, vol. 23, no. 4, pp. 731–757, 1993. https://doi.org/10.1175/1520-0485(1993)023<0731:MCOTHL>2.0.CO;2

v K. Fraedrich, "A suite of user-friendly global climate models: Hysteresis experiments", Eur. Phys. J. Plus, vol. 127, no. 5, p. 53, 2012, https://doi.org/10.1140/epjp/i2012-12053-7

vi K. Fraedrich, et al. "Portable university model of the atmosphere". DKRZ Rep 16, 1998. https://mms.dkrz.de/pdf/reports/ReportNo.16.pdf

vii K. Hasselmann, "An ocean model for climate variability studies", Progress in Oceanography, vol. 11, no. 2, pp. 69–92, 1982, https://doi.org/10.1016/0079-6611(82)90004-0

viii F. Lunkeit, et al. „Planet Simulator Reference Manual", Technical Report, University of Hamburg, 2011, https://www.mi.uni-hamburg.de/en/arbeitsgruppen/theoretische-meteorologie/modelle/sources/psreferencemanual-1.pdf

ix A. J. Semtner, "A Model for the Thermodynamic Growth of Sea Ice in Numerical Investigations of Climate", Journal of Physical Oceanography, vol. 6, no. 3, pp. 379–389, 1976. https://doi.org/10.1175/1520-0485(1976)006<0379:AMFTTG>2.0.CO;2

x P. Paiewonsky et al., "Description and validation of the Simple, Efficient, Dynamic, Global, Ecological Simulator (SEDGES v.1.0)", Geoscientific Model Development, vol. 11, no. 3, pp. 861–901, 2018, https://doi.org/10.5194/gmd-11-861-2018

xi W. A. Müller et al., "The ICON-based Earth System Model for Climate Predictions and Projections (ICON XPP v1.0)", EGUsphere, pp. 1–60, 2025, https://doi.org/10.5194/egusphere-2025-2473

xii O. Mehling et al.,"Parameterization dependence of the hydrological cycle in a general circulation model of intermediate complexity", EGU General Assembly, 2021, https://doi.org/10.5194/egusphere-egu21-1328

xiii F. Pollak et al., "Climate Sensitivity and Convective Parameterization in the Earth System Model of Intermediate Complexity PlaSim", Proceedings of the 8th bwHPC Symposium 2022, 2022.

xiv E. J. Rohling et al., "Making sense of palaeoclimate sensitivity", Nature, vol. 491, no. 7426, pp. 683–691, 2012, https://doi.org/10.1038/nature11574

xv S. C. Sherwood et al., "An Assessment of Earth's Climate Sensitivity Using Multiple Lines of Evidence", Reviews of Geophysics, vol. 58, no. 4, 2020, https://doi.org/10.1029/2019RG000678

xvi K. Rehfeld et al. "Global patterns of declining temperature variability from the Last Glacial Maximum to the Holocene", Nature, 2018, https://doi.org/10.1038/nature25454

xvii M. S. Williamson et al., "Theoretical foundations of emergent constraints: relationships between climate sensitivity and global temperature variability in conceptual models", Dyn Stat Clim Syst, vol. 3, no. 1, 2018, https://doi.org/10.1093/climsys/dzy006

xviii G. Roe, "Feedbacks, Timescales, and Seeing Red", Annual Review of Earth and Planetary Sciences, vol. 37, no. Volume 37, 2009, pp. 93–115, 2009, https://doi.org/10.1146/annurev.earth.061008.134734

xix J. D. Annan et al., "A new global surface temperature reconstruction for the Last Glacial Maximum", Climate of the Past, vol. 18, no. 8, 2022, https://doi.org/10.5194/cp-18-1883-2022

xx X. Yin et al "NOAA Global Surface Temperature Dataset (NOAAGlobalTemp), Version 6.0." NOAA, 2024, https://doi.org/10.1175/BAMS-D-24-0012.1

xxi O. Mehling et al „Parameterization dependence of the hydrological cycle in a general circulation model of intermediate complexity", EGU General Assembly 2021, 2021, https://doi.org/10.5194/egusphere-egu21-1328

xxii M. Adam et al., "The role of dynamic sea ice in a simplified general circulation model used for palaeoclimate studies", in VI ECCOMAS Young Investigators Conference, 2021, https://doi.org/10.4995/YIC2021.2021.12383

The 11th bwHPC Symposium highlights advances in high-performance computing across disciplines in Baden-Württemberg and beyond. Contributions cover research data management, FAIR integration, scalable workflows, and HPC-enabled methods for modeling electrochemical and quantum systems.

The proceedings showcase large-scale simulations, including PET/CT modeling, astrophysical radiative transfer, CFD in agriculture, and N-body studies of supermassive black holes. Climate modeling advances include PaleoPlaSim 1.0. Emerging topics address AI-driven HPC, security challenges, and modern infrastructure needs, while bwRSE4HPC emphasizes sustainable software engineering.

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

TÜBINGEN LIBRARY PUBLISHING

ISBN 978-3-98944-050-0