

JAKOB LUETTGAU

(+33) 07 81 70 65 81 - jakob.luettgau@inria.fr

National Institute for Research in Digital Science and Technology (Inria)

Centre Inria de l'Université de Rennes, France

EDUCATION

Universität Hamburg (UHH), Germany 2017 - 2021

Dr. rer. nat. (Computer Science) magna cum laude

Dissertation: Decision Support for Workflow-Aware High-Performance Storage Systems

Supervisors: Thomas Ludwig, Philip Carns, Michael Kuhn

Universität Hamburg (UHH) 2014 - 2016

MSc (Computer Science)

Thesis: Modeling and Simulation of Tape Libraries for Hierarchical Storage Management Systems

Universität Hamburg (UHH) 2009 - 2014

BSc (Computer Science)

Thesis: Flexible Event Imitation Engine for Parallel Workloads

RESEARCH INTERESTS

Scientific Workflows, High-Performance Storage Systems, Performance Engineering and Visualization, High-Performance Computing, Scientific Software Development, Cyberinfrastructure at Scale, Machine Learning for System Optimization.

PROFESSIONAL EXPERIENCE

Centre Inria de l'Université de Rennes October 2023 - present

Researcher at National Institute for Research in Digital Science and Technology (Inria)

- Starting Researcher Position in KerData Team

University of Tennessee (UTK), Knoxville Sept 2022 - Aug 2023

Research Assistant Professor at Global Computing Laboratory, Dr. Michela Taufer

University of Tennessee (UTK), Knoxville Oct 2021 - Sept 2022

Post-doctoral Research Associate at Global Computing Laboratory, Dr. Michela Taufer

HelmholtzAI and German Climate Computing Center (DKRZ) 2020 - Oct 2021

Helmholtz AI Consultant Team: Earth and Environment

- DKRZ Application Department
- Artificial Intelligence Innovation for Earth System Analytics and Modelling (AIM) Project

Argonne National Laboratory (ANL) July 2019 - Oct 2019

Research Associate

- Analysis of High Performance Storage Systems for In Situ Workloads

German Climate Computing Center (DKRZ) 2019

DKRZ/University of Hamburg Joint Research Group "Scientific Computing"

- Research on HPC I/O and Optimizing Scientific Workflows

Argonne National Laboratory (ANL)*May 2018 - Aug 2018**Research Associate, Analysis of High Performance Storage for Scientific Workflows***German Climate Computing Center (DKRZ)***2016 - 2019**Researcher for the Centre of Excellence for Weather and Climate in Europe (ESiWACE) Project*

- Research and middleware development for EU H2020 project

Universität Hamburg (UHH)*2013 - 2016**Student and teaching assistant in the working group “Scientific Computing”*

- Participation and Development in Research Projects & Teaching

Freelance Work*2012 - 2014**Freelance Work Klest InterNET Ltd.; Hochschullotsen KG; Universitätskolleg*

- Activities: Linux server administration, infrastructure support for a blog farm, a survey tool and traffic monitoring in different projects.

Freelance Work*2010 - 2013**Behörde für Schule und Berufsbildung, Hamburg; Landesinstitut für Lehrerbildung und Schulentwicklung, Hamburg; Literaturzentrum e.V. Hamburg*

- Activities: Development of various online platforms, database management, and websites.

RESEARCH FUNDING

Collaborative Research: SHF: Small: Model-driven Design and Optimization of Dataflows for Scientific Applications*Project Period: 10/01/2023 - 09/30/2025*

- Source of Support: National Science Foundation #2331152, USA
- Total amount: \$ 624,000 (\$ 424,000 at University of Tennessee Knoxville)
- Role: **Co-PI**, with Michela Taufer (PI) and Jack Marquez (Co-PI),
- Location: University of Tennessee Knoxville, USA / Collaborative research with Ewa Deelman, University of South California
- Description: This project has four main research components. First, the project defines a taxonomy of common dataflow motifs used in scientific domains, ranging from simple producer-consumer pairs to complex pipelines with multiple producers and consumers, by mapping these motifs to real scientific applications. Second, the project designs a middleware layer to handle dataflow pipelines executing on HPC, cloud, and edge resources. Third, the project develops a 2-step model for mitigating pipelines that result in data loss and inefficiencies associated with the slowdown in data production or consumption in dataflow pipelines. Finally, the project trains a broader community to utilize the taxonomy, middleware, and model to optimize real scientific applications by identifying potential bottlenecks and making necessary adjustments to maximize pipeline efficiency and accuracy, continuously monitoring and optimizing pipelines to ensure the highest quality scientific output possible.

Discover ACCESS Award: OAC Piloting the National Science Data Fabric: A Platform Agnostic Testbed for Democratizing Data Delivery*Januar 2023 - January 2024*

- Source of Support: Advanced Cyberinfrastructure Coordination Ecosystem: Services & Support Cyberinfrastructure (ACCESS-CI CIS210128), USA
- Total amount: \$ 76,181.14
- Role: **PI**, with Michela Taufer (Co-PI) and Valerio Pascucci (Co-PI)
- Location: University of Tennessee Knoxville, USA

- Description: This project explores data movements between geographically distributed cyberinfrastructure and modes of HPC & cloud convergence. The research explores network latency and throughput over time between different cyberinfrastructure providers and works closely with applications teams to understand opportunities and limitations with new modes of provisioning compute and storage resources to scientists, educators and students.

XSEDE-Startup Allocation Award: Piloting the National Science Data Fabric: A Platform Agnostic Testbed for Democratizing Data Delivery *December 2021 - December 2022*

- Source of Support: Extreme Science and Engineering Discovery Environment (XSEDE CIS-210128)
- Total amount: \$ 185,165.36
- Role: **PI**, with Michela Taufer (Co-PI) and Valerio Pascucci (Co-PI)
- Location: University of Tennessee Knoxville, USA
- Description: This project explores modes of HPC and Cloud convergence by developing middleware and services, as well as porting applications and workflows in collaboration with science teams to work in academic cloud environments such as Jetstream, Open Science Grid, Open Storage Network, as well as established HPC sites such as operated by TACC.

RESEARCH PROJECTS AND PRODUCTS

NSDF: National Science Data Fabric - University of Utah, University of Tennessee, University of Michigan, San Diego Supercomputing Center, Johns Hopkins University *October 2021 - Present*

- Development of cyberinfrastructure to democratize data delivery
- Architecture design of novel Cyberinfrastructure for geo-distributed scientific workflows
- Development of best practices and training for application teams
- Cataloguezation and indexing of public global research data for discoverability in accordance to FAIR data principles

Flux/Fluence: Enabling workflows in converged environments spanning HPC and Cloud - Lawrence Livermore National Laboratory, IBM/RedHead *August 2022 - Present*

- Port, analyse and optimize MPI-workflows on top of Kubernetes Environments
- Exploration of requirements to enable I/O-awareness in resource orchestration and scheduling systems

DYAD: Enabling In-Situ Storage System-Bypass for Legacy Applications - Lawrence Livermore National Laboratory *February 2022 - Present*

- Enabling automatic storage system by-pass for legacy applications (POSIX, STDIO) using FLUX
- Leveraging of fast transport middleware such as UCX
- Exploring optimization opportunities of combining with hierarchical scheduling approaches such as FLUX

Hatchet/Thicket Performance Analysis Tools - Lawrence Livermore National Laboratory *August 2019 - Present*

- Development methodologies to query into call-tree performance data
- Enable context-capture of execution environments as well as build parameters and runtime parameters
- Automation and development of building-blocks to enable common choice-based optimization (e.g., compilers, library variants)

GPU-accelerated Deduplication and Optimization of Application Checkpointing - Argonne National Laboratory, Los Alamos National Laboratory *October 2021 - Present*

- Accelerate deduplication for large-scale checkpointing applications by leveraging Merkle-Trees and Kokkos for parallelization of hash-calculations for GPU-resident application data

Benchmarking and Optimizing I/O Performance of Containerized Workflows through FUSE Object Storage Mappers - Sandia National Laboratories, IBM Cloud *Oktober 2021 - Present*

- Benchmarking and Performance

Flux: Modelling and Simulation of Hierarchical Scheduling Strategies - Lawrence Livermore National Laboratory *August 2021 - 2022*

- Modeling and Simulation of Analytical Performance Models for Hierarchical Scheduling
- Validation of Hierarchical Scheduling Strategies on HPC resources

PyDarshan: Empowering next-generation I/O Performance Analysis - Argonne National Laboratory *August 2019 - Present*

- Development of advanced analysis capability of I/O performance data
- Development of interactive visualization capabilities
- Enabling modern data science and ML approaches for I/O performance analysis

AIM / Helmholtz AI: Machine Learning Consultancy for Researchers in the Earth and Environmental Sciences - German Climate Computing Center (DKRZ), HelmholtzAI, Helmholtz Centre HEREON, Juelich Supercomputing Center (FZJ) *August 2020 - 2021*

- Support application teams across Germany to adopt, setup and optimize machine learning approaches into their workflows

ESiWACE: Centre of Excellence for Weather and Climate in Europe - German Climate Computing Center (DKRZ), University of Reading, Seagate *August 2016 - 2019*

- Development of research prototypes to improve I/O performance for climate and weather models to run at square kilometer resolutions
- Improve and leverage middleware such as NetCDF and HDF5 for optimized I/O performance
- Simulation and Modelling of Hierarchical Storage Systems with a particular focus on tape

The SIOX Project: Scalable I/O for Extreme Performance - German Climate Computing Center (DKRZ), High-Performance Computing Center Stuttgart (HLRS), University of Hamburg (UHH), Technische Universität Dresden / Center for Information Services and High-Performance Computing (ZIH) *2013 - 2014*

- Development of scriptable parallel benchmarks for I/O kernels
- Development of transparent instrumentation libraries and runtimes for C/C++ applications

PUBLICATIONS

Journal Articles

- [1] N. Zhou, G. Scorzelli, **J. Luetzgau**, R. R. Kancharla, J. J. Kane, R. Wheeler, B. P. Croom, P. Newell, V. Pascucci, and M. Taufer. “Orchestration of Materials Science Workflows for Heterogeneous Resources at Large Scale”. In: *The International Journal of High Performance Computing Applications* 37.3-4 (July 2023), pp. 260–271.

- [2] J. Plehn, A. Fuchs, M. Kuhn, **J. Lüttgau**, and T. Ludwig. “Data-Aware Compression for HPC Using Machine Learning”. In: *ACM SIGOPS Operating Systems Review* 56.; Issue 1 (June 2022), pp. 62–69.
- [3] H. Paasche, M. Gross, **J. Lüttgau**, D. S. Greenberg, and T. Weigel. “To the Brave Scientists: Aren’t We Strong Enough to Stand (and Profit from) Uncertainty in Earth System Measurement and Modelling?”. In: *Geoscience Data Journal* (Sept. 2021).
- [4] K. Duwe, **J. Lüttgau**, G. Mania, J. Squar, A. Fuchs, M. Kuhn, E. Betke, and T. Ludwig. “State of the Art and Future Trends in Data Reduction for High-Performance Computing”. In: *Supercomputing Frontiers and Innovations* 7.1 (Apr. 2020), pp. 4–36.
- [5] **J. Lüttgau**, M. Kuhn, K. Duwe, Y. Alforov, E. Betke, J. Kunkel, and T. Ludwig. “Survey of Storage Systems for High-Performance Computing”. In: *Supercomputing Frontiers and Innovations* 5.1 (Apr. 2018), pp. 31–58.
- [6] J. Kunkel and **J. Lüttgau**. “Interaktiver C-Programmierkurs, ICP”. In: *Synergie, Fachmagazin für Digitalisierung in der Lehre* 2 (Nov. 2016), pp. 74–75.

Conference and Workshop Papers

- [1] S. Brink, M. McKinsey, D. Boehme, W. D. Hawkins, C. Scully-Allison, I. Lumsden, T. Burgess, V. Lama, K. E. Isaacs, **J. Luettgau**, Michela Taufer, and O. Pearce. “Thicket: Seeing the Performance Experiment Forest for the Individual Run Trees”. In: *Proceedings of the 32nd International ACM Symposium on High-Performance Parallel and Distributed Computing (HPDC)*. Orlando, Florida, USA: ACM, June 2023, pp. 1–10.
- [2] **J. Luettgau**, H. Martinez, P. Olaya, G. Scorzelli, G. Tarcea, J. Lofstead, C. R. Kirkpatrick, V. Pascucci, and M. Taufer. “NSDF-Services: Integrating Networking, Storage, and Computing Services into a Testbed for Democratization of Data Delivery”. In: *IEEE ACM International Conference on Utility and Cloud Computing (UCC2023)*. Taormina (Messina), Italy, Dec. 2023.
- [3] **J. Luettgau**, G. Scorzelli, V. Pascucci, and M. Taufer. “Development of Large-Scale Scientific Cyberinfrastructure and the Growing Opportunity to Democratize Access to Platforms and Data”. In: *Distributed, Ambient and Pervasive Interactions*. Ed. by N. A. Streitz and S. Konomi. Lecture Notes in Computer Science. Cham: Springer Nature Switzerland, 2023, pp. 378–389.
- [4] **J. Luettgau**, S. Snyder, T. Reddy, N. Awtrey, K. Harms, J. L. Bez, R. Wang, R. Latham, and P. Carns. “Enabling Agile Analysis of I/O Performance Data with PyDarshan”. In: *Proceedings of the SC ’23 Workshops of The International Conference on High Performance Computing, Network, Storage, and Analysis*. SC-W ’23. New York, NY, USA: Association for Computing Machinery, Nov. 2023, pp. 1380–1391.
- [5] P. Olaya, **J. Luettgau**, C. Roa, R. Llamas, R. Vargas, S. Wen, I.-H. Chung, S. Seelam, Y. Park, and Michela Taufer. “Enabling Scalability in the Cloud for Scientific Workflows: An Earth Science Use Case”. In: *Proceedings of IEEE CLOUD*. Chicago, IL, USA: IEEE Computer Society, June 2023, pp. 1–10.
- [6] N. Tan, **J. Luettgau**, J. Marquez, K. Teranishi, N. Morales, S. Bhowmick, F. Cappello, M. Taufer, and B. Nicolae. “Scalable Incremental Checkpointing Using GPU-Accelerated De-Duplication”. In: *Proceedings of the 52nd International Conference on Parallel Processing*. ICPP ’23. New York, NY, USA: Association for Computing Machinery, Sept. 2023, pp. 665–674.
- [7] N. Tan, B. Nicolae, **J. Luettgau**, J. Marquez, K. Teranishi, N. Morales, S. Bhowmick, M. Taufer, and Franck Cappello. “Scalable Checkpointing of Applications with Sparsely Updated Data”. In: *Proceedings of the 52nd International Conference on Parallel Processing (ICPP)*. Salt Lake City, UT, USA: ACM, Aug. 2023, pp. 1–10.
- [8] **J. Luettgau**, S. Caino-Lores, K. Suarez, and M. Taufer. “Reproducing and Extending Analytical Performance Models of Generalized Hierarchical Scheduling”. In: *2nd Workshop on Reproducible Workflows, Data Management, and Security*. Salt Lake City, UT, USA, 2022, p. 6.
- [9] **J. Luettgau**, G. Scorzelli, V. Pascucci, G. Tarcea, C. R. Kirkpatrick, and M. Taufer. “NSDF-Catalog: Lightweight Indexing Service for Democratizing Data Delivery”. In: *2022 IEEE/ACM 15th*

International Conference on Utility and Cloud Computing (UCC). Portland, Oregon, USA, Dec. 2022, pp. 1–10.

- [10] I. Lumsden, **J. Lüttgau**, V. Lama, C. Scully-Allison, S. Brink, K. E. Isaacs, O. Pearce, and M. Taufer. “Enabling Call Path Querying in Hatchet to Identify Performance Bottlenecks in Scientific Applications”. In: *2022 IEEE 18th International Conference on E-Science (e-Science)*. Oct. 2022, pp. 256–266.
- [11] J. Plehn, A. Fuchs, M. Kuhn, **J. Lüttgau**, and T. Ludwig. “Data-Aware Compression for HPC Using Machine Learning”. In: *Proceedings of the Workshop on Challenges and Opportunities of Efficient and Performant Storage Systems*. CHEOPS ’22. New York, NY, USA: Association for Computing Machinery, Apr. 2022, pp. 8–15.
- [12] **J. Lüttgau** and J. Kunkel. “Cost and Performance Modeling for Earth System Data Management and Beyond”. In: *High Performance Computing*. Ed. by R. Yokota, M. Weiland, J. Shalf, and S. Alam. Lecture Notes in Computer Science. Cham: Springer Int. Publishing, 2018, pp. 23–35.
- [13] **J. Lüttgau**, S. Snyder, P. Carns, J. M. Wozniak, J. Kunkel, and T. Ludwig. “Toward Understanding I/O Behavior in HPC Workflows”. In: *2018 IEEE/ACM 3rd International Workshop on Parallel Data Storage & Data Intensive Scalable Computing Systems (PDSW-DISCS)*. Dallas, TX, USA, Nov. 2018, pp. 64–75.
- [14] **J. Lüttgau** and J. Kunkel. “Simulation of Hierarchical Storage Systems for TCO and QoS”. In: *High Performance Computing*. Ed. by J. M. Kunkel, R. Yokota, M. Taufer, and J. Shalf. Lecture Notes in Computer Science. Cham: Springer International Publishing, 2017, pp. 132–144.
- [15] J. M. Kunkel, M. Zimmer, N. Hübbe, A. Aguilera, H. Mickler, X. Wang, A. Chut, T. Bönisch, **J. Lüttgau**, R. Michel, and J. Weging. “The SIOX Architecture – Coupling Automatic Monitoring and Optimization of Parallel I/O”. In: *Supercomputing*. Ed. by J. M. Kunkel, T. Ludwig, and H. W. Meuer. Lecture Notes in Computer Science. Cham: Springer International Publishing, 2014, pp. 245–260.
- [16] **J. Lüttgau** and J. M. Kunkel. “Feign: In-Silico Laboratory for Researching I/O Strategies”. In: *2014 9th Parallel Data Storage Workshop*. New Orleans, LA, USA, Nov. 2014, pp. 43–48.

Reports

- [1] **J. Lüttgau**, J. Kunkel, B. Lawrence, J. Jensen, G. Congiu, H. Hua, and N. Paola. *New Storage Layout for Earth System Data (D4.2) (Version 2.0)*. Tech. rep. Centre of Excellence for Weather and Climate in Europe (ESiWACE), Nov. 2019.
- [2] J. Kunkel and **J. Lüttgau**. *Prototypes of Alternative Storage Backends (MS7)*. Tech. rep. Centre of Excellence for Weather and Climate in Europe (ESiWACE), May 2018.
- [3] Y. Alforov, E. Betke, K. Chasapis, A. Fuchs, F. Große, N. Jumah, M. Kuhn, J. Kunkel, H. Lenhart, **J. Lüttgau**, P. Neumann, A. Novikova, J. Squar, and T. Ludwig. *Wissenschaftliches Rechnen - Scientific Computing - 2016*. Tech. rep. Deutsches Klimarechenzentrum GmbH, Hamburg, Germany: Research Group: Scientific Computing, University of Hamburg, June 2017.
- [4] J. Kunkel, **J. Lüttgau**, and E. Betke. *Operational Demonstrator of ESD Middleware (MS5)*. Tech. rep. Centre of Excellence for Weather and Climate in Europe (ESiWACE), Aug. 2017.
- [5] **J. Lüttgau**, J. Kunkel, J. Jensen, and B. Lawrence. *Business Model with Alternative Scenarios (D4.1)*. Tech. rep. Centre of Excellence for Weather and Climate in Europe (ESiWACE), Feb. 2017.
- [6] **J. Lüttgau**, J. Kunkel, B. Lawrence, J. Jensen, G. Congiu, H. Hua, and P. Nassisi. *New Storage Layout for Earth System Data (D4.2)*. Tech. rep. Centre of Excellence for Weather and Climate in Europe (ESiWACE), July 2017.

Posters, Abstracts and Short-Papers

- [1] **J. Lüttgau**, H. Martinez, G. Tarcea, G. Scorzelli, V. Pascucci, and M. Taufer. “Studying Latency and Throughput Constraints for Geo-Distributed Data in the National Science Data Fabric”. In: *Proceedings of the 32nd International Symposium on High-Performance Parallel and Distributed*

- Computing*. HPDC '23. New York, NY, USA: Association for Computing Machinery, Aug. 2023, pp. 325–326.
- [2] **J. Luettgau**, P. Olaya, N. Zhou, G. Scorzelli, V. Pascucci, and M. Taufer. “NSDF-Cloud: Enabling Ad-Hoc Compute Clusters Across Academic and Commercial Clouds”. In: *Proceedings of the 31st International Symposium on High-Performance Parallel and Distributed Computing*. Minneapolis MN USA: ACM, June 2022, pp. 279–280.
 - [3] P. Olaya, **J. Luettgau**, N. Zhou, J. Lofstead, G. Scorzelli, V. Pascucci, and M. Taufer. “NSDF-FUSE: A Testbed for Studying Object Storage via FUSE File Systems”. In: *Proceedings of the 31st International Symposium on High-Performance Parallel and Distributed Computing*. HPDC '22. New York, NY, USA: Association for Computing Machinery, June 2022, pp. 277–278.
 - [4] J.-S. Yeom, D. H. Ahn, I. Lumsden, **J. Luettgau**, S. Caino-Lores, and Taufer, Michela. “Ubique: A New Model for Untangling Inter-Task Data Dependence in Complex HPC Workflows”. In: *Proceedings of the 18th IEEE International Conference on E-Science (eScience)*. Salt Lake City, Utah, USA: IEEE Computer Society, Oct. 2022, pp. 1–2.
 - [5] **J. Lüttgau**, T. Weigel, and D. Greenberg. “Data Challenges for Climate Machine Learning Workflows”. In: *GEOMAR 5th Data Science Symposium 2021*. Vol. 5. 2021.
 - [0] T. Weigel, **J. Luettgau**, C. Kadow, D. Greenberg, L. Bouwer, C. Schrum, and T. Ludwig. “AI for Earth System Sciences: Bottlenecks, Pitfalls and Recommendations”. In: *Platform for Advanced Scientific Computing (PASC21)*. Geneva, July 2021.
 - [6] T. Weigel, **J. Lüttgau**, C. Kadow, D. Greenberg, L. M. Bouwer, H. Bockelmann, C. Schrum, and T. Ludwig. “AI for Earth System Sciences: Bottlenecks, Pitfalls and Recommendations”. In: PASC21, July 2021.
 - [7] D. Uhle, M. Hort, **J. Lüttgau**, and J. Walda. “Real-Time Detections of Volcanic Eruptions Using Neuronal Networks”. In: *79. Jahrestagung Deutsche Geophysikalische Gesellschaft*. Braunschweig, Germany, Mar. 2019.
 - [8] **J. Lüttgau**, E. Betke, O. Perevalova, J. Kunkel, and M. Kuhn. “Adaptive Tier Selection for NetCDF and HDF5”. In: *Supercomputing Conference 2017 (SC17)*. Denver, CO, USA, Nov. 2017.
 - [9] **J. Lüttgau**, J. Kunkel, B. N. Lawrence, S. Fiore, and H. Hua. “Towards Structure-Aware Earth System Data Management”. In: 2nd Joint International Workshop on Parallel Data Storage & Data Intensive Scalable Computing Systems (PDSW-DISCS) WIPs, 2017.
 - [10] J. Kunkel, **J. Luettgau**, B. N. Lawrence, J. Jensen, G. Congiu, and J. Readey. “Middleware for Earth System Data”. In: 1st Joint International Workshop on Parallel Data Storage & Data Intensive Scalable Computing Systems (PDSW-DISCS) WIPs, 2016.
 - [11] **J. Lüttgau** and J. Kunkel. “Modeling and Simulation of Tape Libraries for Hierarchical Storage Systems”. In: *Supercomputing Computing (SC16)*. Salt Lake City, Utah, USA, Nov. 2016.
 - [12] J. Kunkel, T. Ludwig, **J. Lüttgau**, D. Timmermann, C. Kautz, and V. Skwarek. *Interaktiver C Kurs (ICP)*. Hamburg, Nov. 2015.

Book Chapters

- [1] J. Kunkel and **J. Lüttgau**. “Interaktiver C-Programmierkurs, ICP”. In: *HOOU Content Projekte Der Vorprojektphase 2015/16 – Sonderband Zum Fachmagazin Synergie*. Universität Hamburg, Mittelweg 177, 20148 Hamburg: Universität Hamburg, Apr. 2017, pp. 182–186.

Theses

- [1] **J. Lüttgau**. “Decision Support for Workflow-Aware High-Performance Storage Systems”. Dissertation. Hamburg: Universität Hamburg, May 2021.
- [2] **J. Lüttgau**. “Modeling and Simulation of Tape Libraries for Hierarchical Storage Management Systems”. Master’s Thesis. Hamburg: Universität Hamburg, Sept. 2016.
- [3] **J. Lüttgau**. “Flexible Event Imitation Engine for Parallel Workloads”. Bachelor’s Thesis. Hamburg: Universität Hamburg, Mar. 2014.

TALKS

Conference and Workshop Talks

- [1] *Enabling Agile Analysis of I/O Performance Data with PyDarshan*. New York, NY, USA, Nov. 2023.
- [2] *Data Challenges for Climate Machine Learning Workflows*. GEOMAR 5th Data Science Symposium. (Virtual), 2021.
- [0] *IOBAT: I/O Behavior Analysis Toolkit for HPC Workflows*. Online, Dec. 2021.
- [3] *Cost and Performance Modeling for Earth System Data Management and Beyond*. HPC-IODC: HPC I/O in the Data Center Workshop (ISC-HPC). Frankfurt, Germany, June 2018.
- [4] *Toward Understanding I/O Behavior in HPC Workflows*. 2018 IEEE/ACM 3rd International Workshop on Parallel Data Storage & Data Intensive Scalable Computing Systems (PDSW-DISCS) at SC18. Dallas, TX, USA, Nov. 2018.
- [5] *Simulation of Hierarchical Storage Systems for TCO and QoS*. HPC-IODC: HPC I/O in the Data Center Workshop (ISC-HPC). Frankfurt, Germany, 2017.

Invited Talks

- [1] “NSDF-Catalog: Lightweight Indexing Service for Democratizing Data Delivery”. In: (Apr. 2023).
- [2] *NSDF Software Development Life Cycle Procedures*. Feb. 2022.
- [3] *NSDF-Catalog: Toward a Lightweight Indexing Service for the National Science Data Fabric*. Oct. 2022.
- [4] *Efficient Data Loading and Benchmarking Based on WeatherBench*. HelmholtzAI All-Hands Meeting. (Virtual), June 2021.
- [5] *IOBAT: I/O Behavior Analysis Toolkit for HPC Workflows*. 13th JLESC Workshop. Joint Laboratory for Extreme Scale Computing (JLESC) (Virtual), Dec. 2021.
- [6] *IOBAT: Toward Interactive I/O Analysis for HPC Workflows*. Parallel Computing and I/O (ParCIO) Seminar. Otto von Guericke University Magdeburg (OVGU) (Virtual), May 2021.
- [7] *AIM: AI Innovation for Earth System Analytics and Modelling*. DKRZ Seminar. Germany Climate Computing Center (DKRZ) (Virtual), Oct. 2020.
- [8] *A Peek into Workflow I/O*. SC18, BoF Analyzing Parallel I/O. Dallas, TX, USA, 2018.
- [9] *Augmenting Workflow I/O with Darshan/TOKIO*. Mathematics and Computer Science (MCS) Division Seminar. Argonne National Laboratory, Lemont, IL, USA, July 2018.
- [10] *Parallel I/O & Benchmarking*. Second NESUS Winter School & PhD Symposium 2017. Vibo Valentia, Italy, Feb. 2017.
- [11] *Structure-Aware Adaptive Data Placement*. Leogang High Performance Computing Workshop 2017. Leogang, Austria, Mar. 2017.
- [12] *SIOX In-Situ Optimization and Virtual Laboratory*. SC16 BoF: Analyzing Parallel I/O. Salt Lake City, USA, 2016.
- [13] *SIOX Replay - Laboratory for I/O Research*. CluStor: Workshop on Cluster Storage Technology. German Climate Computing Center (DKRZ), Hamburg, Germany, 2015.

TEACHING

Courses and Seminars

- [1] *Proseminar "Efficient Programming in C" / „Effiziente Programmierung in C“*. University of Hamburg (UHH): Summer-Semester/Sommersemester, 2021.
- [2] *Proseminar "Software Development in the Sciences" / „Softwareentwicklung in Der Wissenschaft“*. University of Hamburg (UHH): Summer-Semester/Sommersemester, 2021.
- [3] *Proseminar "Python for High-Performance Computing" / „Python Im Hochleistungsrechnen“*. University of Hamburg (UHH): Summer-Semester/Sommersemester, 2019.

- [4] *Proseminar "Software Development in the Sciences" / „Softwareentwicklung in Der Wissenschaft“*. University of Hamburg (UHH): Summer-Semester/Sommersemester, 2019.
- [5] *Praktikum "Programming in C" / „C-Programmierung“*. University of Hamburg (UHH): Winter-Semester/Wintersemester, 2018.
- [6] *Projekt "Evaluation of Parallel Computing" / „Parallelrechnerevaluation“*. University of Hamburg (UHH): Winter-Semester/Wintersemester, 2018.
- [7] *Proseminar "Storage and File Systems" / „Speicher- Und Dateisysteme“*. University of Hamburg (UHH): Winter-Semester/Wintersemester, 2018.
- [8] *Seminar "Efficient Programming" / „Effiziente Programmierung“*. University of Hamburg (UHH): Winter-Semester/Wintersemester, 2018.
- [9] *Projekt „Big Data“*. University of Hamburg (UHH): Winter-Semester/Wintersemester, 2017.
- [10] *Proseminar "Storage and File Systems" / „Speicher- Und Dateisysteme“*. University of Hamburg (UHH): Winter-Semester/Wintersemester, 2017.
- [11] *Seminar "Efficient Programming" / „Effiziente Programmierung“*. University of Hamburg (UHH): Winter-Semester/Wintersemester, 2017.
- [12] *Seminar "Latest Trends in Big Data Analytics" / „Neueste Trends in Big Data Analytics“*. University of Hamburg (UHH): Winter-Semester/Wintersemester, 2017.
- [13] *Projekt "Evaluation of Parallel Computing" / „Parallelrechnerevaluation“*. University of Hamburg (UHH): Summer-Semester/Sommersemester, 2016.
- [14] *Proseminar "Programming in R" / „Programmierung in R“*. University of Hamburg (UHH): Summer-Semester/Sommersemester, 2016.

Tutorials and Trainings

- [1] *Machine Learning for Marine Sciences, Part 1: Introduction*. MarDATA Courses for Doctoral Students. HelmholtzAI, Sept. 2021.
- [2] *Machine Learning for Marine Sciences, Part 2: Data Preprocessing*. MarDATA Courses for Doctoral Students. HelmholtzAI, Sept. 2021.

MENTORING

Students at UTK

co-mentored with Michela Taufer

- Andrew Mueller (June 2023 - August 2023)
- Julius Plehn (January 2023 - July 2023)
- Vanessa Lama (March 2022 - May 2023, graduated MSc)
- Ian Lumsden (April 2022 - August 2023)
- Nigel Tan (Oktober 2021 - August 2023)
- Paula Olaya (Oktober 2021 - August 2023)
- Treece Burgess (Oktober 2021 - August 2023)
- Kae Suarez (Oktober 2021 - May 2022, graduated MSc)

Students at UHH

- Julius Plehn (July 2021 - May 2022, graduated MSc)
- Daniel Bremer (Jan 2021 - June 2021, graduated MSc)
- Daniel Uhle (November 2018 - June 2019, graduated MSc)
- Leo Reichenbach (April 2018 - Sept 2018, graduated BSc)
- Roman Jerger (November 2017 - March 2018, graduated MSc)
- Marc Perzborn (March 2016 - Oct 2016, graduated BSc)

Students in cooperation with other Institutions

- Lennart Borchers (Decemeber 2022 - present, OVGU and UHH)
- Enno von Hartmann (March 2022 - Oktober 2022, graduated BSc, OVGU)
- Ravi Mallikarjun Yadav, Chennaboina (March 2022 - June 2022, graduated BSc, OVGU)

Advised Master Theses

- [1] V. Lama. “Interactive Data Analysis of Multi-Run Performance Data”. MA thesis. May 2023.
- [2] J. Plehn. “Data-Aware Compression for HPC Using Machine Learning”. MA thesis. Universität Hamburg, May 2022.
- [3] Ravi Mallikarjun Yadav, Chennaboina. “Domain-Specific Compression Using Auto-Encoders For Climate Data”. MA thesis. Otto von Guericke University Magdeburg (OVGU), May 2022.
- [4] D. Bremer. “Scimon - Scientific Monitor for Automated Run Logging and Reproducibility”. MA thesis. Universität Hamburg, June 2021.
- [5] R. Jerger. “Modeling and Performance Prediction of HDF5 Data on Object Storage”. MA thesis. Universität Hamburg, Mar. 2018.

Advised Bachelor Theses

- [1] L. Reichenbach. “Learned Index Structures: An Evaluation of Their Performance in Key-Value Storage Solutions”. Bachelor’s Thesis. Universität Hamburg, Sept. 2018.
- [2] M. Perzborn. “Evaluation von Alternativen Speicherszenarien Für Hierarchische Speichersysteme”. Bachelor’s Thesis. Universität Hamburg, Oct. 2016.

SERVICES

2024

- EuroSys 2024 Workshop: Challenges and Opportunities of Efficient and Performant Storage Systems (CHEOPS) – **Program Committee**
- Workshop: International Workshop on Converged Computing (WOCC’24) - **Program Committee**
- SC23: Reproducibility Challenge - **Planning Committee**
- Journal for Future Generation Computer Systems – **Reviewer**
- ICPP’24: 53rd International Conference on Parallel Processing - **Program Committee**
- IEEE Computing in Science and Engineering (CiSE) Special Issue: Converged Computing: A best-of-both Worlds of HPC and Cloud - 2024 (Sep/Oct) Issue - **Guest Editor**

2023

- ACM HPDC: The 32st International Symposium on High-Performance Parallel and Distributed Computing- **Program Committee**
- EuroSys 2023 Workshop: Challenges and Opportunities of Efficient and Performant Storage Systems (CHEOPS) – **Program Committee**
- Workshop: International Workshop on Converged Computing (WOCC’23) - **Program Committee**
- 9th International Workshop on Data Analysis and Reduction for Big Scientific Data (DRBSD 2023) - **Program Committee**
- 5th Annual Workshop on Extreme-scale Experiment-in-the-Loop Computing (XLOOP 2023) - **Program Committee**
- PFAIR23: Practically FAIR 2023 Workshop - **Program Committee**
- ISC23 BoF: Ethics in HPC - **Organizer**
- ISC23 BoF: National Science Data Fabric- **Organizer**
- SC23: Reproducibility AD/AE Appendices - **Planning Committee**

- SC23 BoF: Ethics in HPC - **Organizer**
- SC23 BoF: National Science Data Fabric- **Organizer**
- ICPP'23 - **Reproducibility Committee**
- SBAC-PAD 2023 - **Program Committee**
- Journal for Future Generation Computer Systems – **Reviewer**
- IEEE Transactions on Parallel and Distributed Systems – **Reviewer**

2022

- ReWorDS22 Workshop – **Program Committee**
- EuroSys 2022 Workshop: Challenges and Opportunities of Efficient and Performant Storage Systems (CHEOPS) – **Program Committee**
- IEEE Cluster 2022 - **Delegate Reviewer**
- Journal for Future Generation Computer Systems – **Reviewer**
- 1st Workshop on Trustable and Ethical Machine Learning – **Program Committee**
- 4th Annual Workshop on Extreme-scale Experiment-in-the-Loop Computing (XLOOP 2022) - **Program Committee**
- SC22 BoF: Ethics in HPC - **Organizer**

2021

- EuroSys 2021 Workshop: Challenges and Opportunities of Efficient and Performant Storage Systems (CHEOPS) – **Program Committee**
- SC21: Supercomputing – **Delegate Reviewer**
- SC21: BoF: Ethics in HPC – **Organizer**
- ReWorDS21 Workshop – **Program Committee**
- Journal for Future Generation Computer Systems – **Reviewer**

2020

- Euro-Par 2020 Workshop: CHAOSS: Challenges and Opportunities of HPC Storage Systems - **Program Committee**

2019

- SC19 BoF: Ethics in HPC - **Organizer**
- C3DIS 2019: Workshop: Machine Learning for Earth and Space Sciences – **Organizer**

2017

- ISC HPC 2017 - **Delegate Reviewer**

PROFESSIONAL AFFILIATIONS

- ACM, ACM SIGHPC
- IEEE, IEEE Computer Society
- RISC-V International Member