High Performance Computing

Atomistic Simulation | Life Science | Agent-based Simulation

18.09.23 - 20.09.23 | Zuse Institute Berlin

Book of Abstracts

Scientific Committee and Support

Special Thanks To

We would like to thank the Zuse Institute Berlin and the Institute of Computer Science at Freie Universität Berlin for their kind support with preparing and hosting the conference and for providing the conference facilities free of charge.

Scientific Committee

Prof. Dr. Petra Imhof, FAU Erlangen-Nürnberg

Prof. Dr. Thomas Kühne, TU Dresden u. Helmholtz-Zentrum Dresden-Rossendorf

Prof. Dr. Michael Schroeder, TU Dresden

Dr. Sarah Wolf, FU Berlin

Prof. Dr. Felix Wolf, TU Darmstadt

NHR Conference ' 23 supported by CECAM-DE_MMS

Technical Support

Local Support	Zuse Institute Berlin and Institute of Computer Science at Freie Universität Berlin
	Anita Ragyanszki
	Ronaldo Rodrigues Pela
Registration Support	Universität Paderborn - Center for Parallel Computing PC^2
	Carsten Bauer
	Robert Schade

Organizing Team

NHR Office

Sponsors

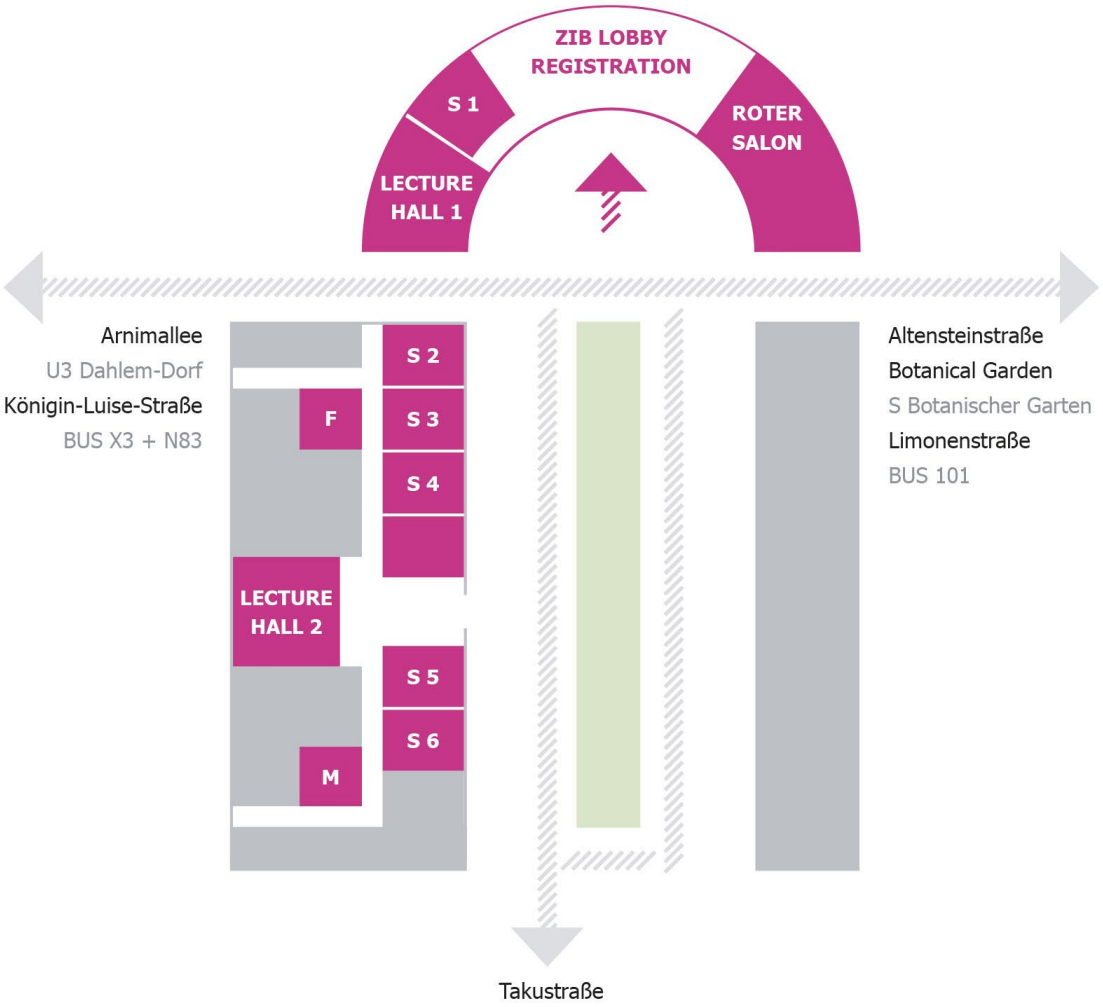
SPONSORED BY THE

And the Federal States participating in the NHR

Location Map

Location

Zuse Institute Berlin (ZIB)
Takustraße 7
14195 Berlin, Germany



Abstract ID - Abbreviation List

KN	Keynote Talk
IN	Invited Talk
CT	Contributed Talk
P	Poster

AS	Atomistic Simulation
LS	Life Science
ABS	Agent-based Simulation
O	Other

Schedule Scientific Conference

Monday, September 18

11:00–11:20	Opening Session Christof Schütte –Opening Remarks Session Chair: Sarah Wolf	Lecture Hall 1
11:20–12:20	Keynote Lecture Agent-based Simulation Robert Axtell "Modeling the U.S. economy, every worker, every firm" [KN-ABS-01] Session Chair: Sarah Wolf	Lecture Hall 1
12:20-13:20	Lunch	ZIB Lobby
13:20-13:50	Invited Talk Atomistic Simulation Claudia Draxl "Electronic-structure theory and materials discovery on the way to exascale"[IN-AS-01] Session Chair: Petra Imhof	Lecture Hall 1
	Contributed Talks Atomistic Simulation	Lecture Hall 1
13:50-14:05	Hermann Lederer [CT-AS-01]	
14:05-14:20	Philipp Neumann [CT-AS-02]	
14:20-14:35	Jadran Vrabec [CT-AS-03]	
14:35-14:50	Martin Brehm [CT-AS-04]	
14:50-15:05	Max Aehle [CT-AS-05] Session Chair: Petra Imhof	
13:20-13:50	Invited Talk Life Science Jens Meiler "High-Performance Computing in Structural Biology and Drug Discovery"[IN-LS-01] Session Chair: Michael Schröder	Lecture Hall 2
	Contributed Talks Life Science	Lecture Hall 2
13:50-14:05	Rainer Böckmann [CT-LS-01]	
14:05-14:20	Mohsen Sadeghi [CT-LS-02]	
14:20-14:35	Yannis Kalaidzidis [CT-LS-03]	
14:35-14:50	Marvin Kaster [CT-LS-04] Session Chair: Michael Schröder	
13:20-13:50	Invited Talk Agent-based Simulation Gary Polhill "High-end computing support for large-scale agent-based modelling – challenges and opportunities"[IN-ABS-01] Session Chair: Sarah Wolf	Seminar Room 1
	Contributed Talks Agent-based Simulation	Seminar Room 1
13:50-14:05	Steffen Fürst [CT-ABS-01]	
14:05-14:20	Jürgen Groeneveld [CT-ABS-02]	
14:20-14:35	Björn Goldenbogen [CT-ABS-03];	
14:35-14:50	N.N. [CT-ABS-04]; Session Chair: Sarah Wolf	

15:05-15:30	Coffee Break	ZIB Lobby
15:30-16:30	Keynote Lecture Atomistic Simulation Helmut Grubmüller "Nanomachines at work: atomistic simulations of macromolecular dynamics and function" [KN-AS-01] Session Chair: Thomas Kühne	Lecture Hall 2
16:30-18:30	Poster Session Finger Food & Beverages	ZIB Lobby + Outdoor

Tuesday, September 19

09:00–10:00	Keynote Lecture Life Science Karissa Sanbonmatsu "Integrative structural biology studies of nucleic acid systems: ribosomes and chromatin " [KN-LS-01] Session Chair: Felix Wolf	Lecture Hall 2
10:00-10:20	Coffee Break	ZIB Lobby
10:20-10:50	Invited Talk Atomistic Simulation Karsten Albe "Accelerating the development of battery materials by high-performance computing "[IN-AS-02] Session Chair: Petra Imhof	Lecture Hall 1
	Contributed Talks Atomistic Simulation	Lecture Hall 1
10:50-11:05	Peter Hildebrand [CT-AS-06]	
11:05-11:20	Jan Kloppenburg [CT-AS-07]	
11:20-11:35	Karan Shah [CT-AS-08]	
11:35-11:50	Linus Erhard [CT-AS-09]	
11:50-12:05	Daniel Fritsch [CT-AS-10] Session Chair: Petra Imhof	
10:20-10:50	Invited Talk Life Science Birgit Strodel "A holistic approach to PET-degrading enzyme design: from QM/MM and MD simulations to ML and experiments " [IN-LS-02] Session Chair: Felix Wolf	Lecture Hall 2
	Contributed Talks Life Science	Lecture Hall 2
10:50-11:05	Julian Herold [CT-LS-05]	
11:05-11:20	Ali Al-Fatlawi [CT-LS-06]	
11:20-11:35	Christian Faber [CT-LS-07]	
11:35-11:50	Ali Doosthosseini [CT-LS-08]	
11:50-12:05	Chris Lauber [CT-LS-09] Session Chair: Felix Wolf	

	Mixed Session: Life Science & Atomistic Simulation	Seminar Room 1
10:20-10:35	James Bowden [CT-LS-10]	
10:35-10:50	Kathrin Luise Braband [CT-LS-11]	
10:50-11:05	Niklas Leimeroth [CT-AS-11]	
11:05-11:20	Philip Langer [CT-LS-12]	
11:20-11:35	Michael Winter [CT-AS-12]	
11:35-11:50	Xiaofei Ping [CT-LS-13]	
	Session Chair: Anita Ragyanszki	
12:05-13:15	Lunch	ZIB Lobby
13:15–14:15	Keynote Lecture Life Science	Lecture Hall 2
	Mohammed AlQuraishi "OpenFold: Lessons learned and insights gained from rebuilding and retraining AlphaFold2" [KN-LS-02] Session Chair: Michael Schröder	
14:15-14:30	Coffee Break	ZIB Lobby
14:30-16:00	Panel Discussion " Women in HPC: Empowering future careers"	Lecture Hall 1
16:00-16:30	Coffee Break	ZIB Lobby
16:30-18:00	Panel Discussion + Closing Session "NHR – HPC for Science"	Lecture Hall 1

Schedule NHR Internal Day

Tuesday, September 19

10:20-11:50	NHR Working Group Security	Seminar Room 2
10:20-11:50	NHR Earth System Sciences	Seminar Room 3
10:20-11:50	NHR Public Relation	Seminar Room 4
13:15-14:30	NHR System Administration	Seminar Room 1

Wednesday, September 20

09:00–10:30	Meeting Nutzungsausschusses – non public	Seminar Room 1
09:00-10:30	Joint Meeting Mitgliederversammlung and Betreiberausschuss – non public	Roter Salon
09:00-10:30	Career Planning Meeting NHR Graduate School Scholars with HPC-Experts	Seminar Room 2
09:00:09:30	Consulting Teams of NHR Alliance (CNHR): Joint part	Lecture Hall 1
09:30-10:30	CNHR Breakout - Life Science	Lecture Hall 1
09:30-10:30	CNHR Breakout – Digital Humanities	Seminar Room 3
09:30-10:30	CNHR Breakout – Physics/ Chemistry	Seminar Room 4
09:30-10:30	CNHR Breakout - CFD	Seminar Room 5
10:30-11:00	Coffee Break	ZIB Lobby
11:00-12:00	Presentation Central Projects/ NHR Working Groups Part I 11:00-11:10 Patrick Gelß: "Quantum Computing and Integer Programming" 11:10-11:20 Thorsten Reimann: "Koordination NHR mit HPC-Ländernetzwerken" 11:20-11:30 Sebastian Döbel: "Integration of Deep Learning APIs into existing performance analysis tools" 11:30-11:40 Freja Nordsiek: "Container und Container Management" 11:40-11:50 René Caspart: "Cx as a service for sustainable HPC research software development" 11:50-12:00 Hendrik Nolte: "Data Lakes"	Lecture Hall 1
11:00-12:00	NHR Office Administration	Seminar Room 1

12:00-13:00	Lunch & Poster Session NHR Graduate School Fellows and Central Projects	ZIB Lobby + Outdoor
13:00-14:00	Presentation Central Projects/ NHR Working Groups Part I	Lecture Hall 1
	13:00-13:10 Thomas Steinke: "Performance Lab 1 – Heterogene Architekturen" 13:10-13:20 Thomas Steinke: "Optimierung von Bibliotheken für datenparallele Prozessorarchitekturen" 13:20-13:30 Frank Winkler: "Interoperabilität + gemeinsame Standards für eine system-weite + kontinuierliche Job-spezifische Performance-Monitoring Umgebung" 13:30-13:40 Fabian Lingenhöl: "NHR Sicherheitskreis" 13:40-13:50 Georg Hager: "AG Training" 13:50-14:00 Tim Cramer: "NHR Resource Allocation with JARDS"	
14:00-14:15	Coffee Break	ZIB Lobby
14:15-15:30	Resume NHR	Lecture Hall 1
15:30-19:00	NHR Team Event	ZIB Outdoor

Abstracts Keynote Lectures

Abstract ID: KN-AS-01

Title: Nanomachines at work: atomistic simulations of macromolecular dynamics and function

Speaker: Helmut Grubmüller, Max Planck Institute for Multidisciplinary Sciences Göttingen

Co-Authors: Maxim Igaev, Lars V. Bock, Leonard Heinz, Steffen Schultze

We will describe four routes towards revealing molecular dynamics and function, combining high performance atomistic simulations, cryo-electron microscopy, statistical mechanics, and Bayesian approaches. (1) We used non-equilibrium atomistic simulations of whole microtubules, comprising ca. 25 Mio particles, to address the question how they switch between growing and shrinking phase, despite their very similar structure of the tip. Our results show that and why the primary steps of microtubule tip flaring differ kinetically between GTP and GDP loaded states. (2) For improved ensemble refinement from cryo-electron microscopy data, we used continuum and atomistic simulations of shock freezing of whole fully solvated ribosomes to quantify how much of the physiological temperature structural heterogeneity and dynamics is preserved in electron microscopy samples at cryogenic temperatures. (3) The thermodynamics of solvent shells determines biomolecular structures, yet to spatially resolve and quantify the tug of war between enthalpy and entropy is still a considerable challenge. Using permutation symmetry and mutual information expansions, we show for the example protein crambin that more than half of the solvent entropy contribution arises from induced water-water correlations. This unexpected finding seems to contradict the well-known Ben-Naim theorem -- but wait. (4) We will address structure refinement from single molecule femtosecond XFEL diffraction images, for which the signal to noise is in the extreme Poisson regime. Using synthetic data we have demonstrated via a ratationally invariant correlation approach that near-atomistic resolution is possible even for small proteins. With a rigorous Bayesian approach, even ensemble refinement can be achieved with much fewer images than expected from an information theoretical perspective.

Abstract ID: KN-ABS-01

Title: Modeling the U.S. economy, every worker, every firm

Speaker: Robert Axtell, George Mason University Fairfax

The availability of firm-level micro-data for the U.S. private sector has given unprecedented insight into the empirical structure of American business firms. Within these data are dozens of previously unknown patterns and gross regularities, few of which can be readily explained by appealing to conventional economic theories. Such theories typically utilize a single representative agent whose actions are rational and occur in equilibrium. I will describe a family of so-called agent-based models, realized computationally for 120 million worker agents who self-organize into 6 million groups having many of the properties of U.S. businesses. Specifically, both real and modeled firms have power law distributions of size, Weibull distributions of age, Laplace distributions of log growth rates, heavy-tailed productivity distributions, and are spatially arranged in clusters (e.g., cities) whose sizes are also heavy-tailed. Additionally, a variety of networks, governing the flow of goods and labor, are described. This code is naturally parallelized and several different approaches to doing so will be described. This work is an example of a new kind of economic model that uses agent heterogeneity, bounded rationality, networks, and out-of-equilibirum behavior to create larger-scale and more realistic models than those typically used by central banks, work made possible by the confluence of firm-level micro-data and parallel computing hardware.

Abstract ID: KN-LS-01

Title: Integrative structural biology studies of nucleic acid systems: ribosomes and chromatin

Speaker: Karissa Sanbonmatsu, Los Alamos National Laboratory, Los Alamos

TBA

Abstract ID: KN-LS-02

Title: OpenFold: Lessons learned and insights gained from rebuilding and retraining AlphaFold2

Speaker: Mohammed AlQuraishi, Columbia University New York

AlphaFold2 revolutionized structural biology by accurately predicting protein structures from sequence. Its implementation however (i) lacks the code and data required to train models for new tasks, such as predicting alternate protein conformations or antibody structures, (ii) is unoptimized for commercially available computing hardware, making large-scale prediction campaigns impractical, and (iii) remains poorly understood with respect to how training data and regimen influence accuracy. Here we report OpenFold, an optimized and trainable version of AlphaFold2. We train OpenFold from scratch and demonstrate that it fully reproduces AlphaFold2's accuracy. By analyzing OpenFold training, we find new relationships between data size/diversity and prediction accuracy and gain insights into how OpenFold learns to fold proteins during its training process.

Abstracts Invited Talks

Abstract ID: IN-LS-01

Title: High-Performance Computing in Structural Biology and Drug Discovery

Speaker: Jens Meiler, Universität Leipzig

I will review recent developments in computational structural biology and drug discovery including protein structure prediction with AlphaFold, protein design with ProteinMPNN, and drug discovery via ultra large library screening

Abstract ID: IN-AS-01

Title: Electronic-structure theory and materials discovery on the way to exascale

Speaker: Claudia Draxl, Humboldt-Universität zu Berlin

The discovery of improved or even novel materials with desired properties is one of the most important applications of high-performance computing (HPC). Impressive progress has been and is being made in the field of quantum-mechanics-based *ab initio* computational materials science to describe and understand the various competing interactions that take place at the atomistic and electronic scale. However, the application of these methods to complex materials is often hampered by the enormous demand in terms of computing power, especially when high accuracy is required. The advent of exascale computing promises accelerated insight into the nature of materials and the discovery of new ones. However, it comes with several challenges. These include dramatically accelerating community codes to optimally scale up to hundreds of thousands of cores, high-throughput calculations of large fractions of the (in principle infinite) materials space, and handling extreme-scale data with novel artificial-intelligence tools. In this talk, I will address current limitations and present algorithms that have allowed us to overcome some of the bottlenecks and achieve benchmark quality with currently available computing power. I will also discuss how high-level approaches can be brought to the exascale as currently pursued in the European Center of Excellence NOMAD (https://nomad-coe.eu).

Abstract ID: IN-AS-01

Title: High-end computing support for large-scale agent-based modelling – challenges and opportunities

Speaker: Gary Polhill, The James Hutton Institute Aberdeen

One of the earliest agent-based models was reputedly first simulated by hand on a checkerboard. Over the more than half a century since, both computing power and agent-based modelling have experienced huge growth. From early thought experiments using computer simulation, agent-based modelling has increasingly been applied in empirical contexts, and rose to prominence during the Covid crisis. Though there are a few exceptions, most people developing and using agent-based models still run them on their personal computer, meaning that the community has not kept up with the last 20 years of advances in high-end computing. That's a pity, because empirical agent-based modellers could do a much better job of applying their approach to real-world situations if they took full advantage of the computing power now available to run larger scale models and make more thorough searches of parameter spaces and configuration options. With the advent of exascale computing, it's time to think about how we address the institutional and technical obstacles to more everyday use of high-end computing by the agent-based modelling community. This talk will explain the specialist requirements of the community for high-end computing, outline the bureaucratic and managerialist practices that systematically obstruct this community from accessing it, and suggest ways we might work around these issues that would doubtless be of wider benefit to all high-end computing users.

Abstract ID: IN-LS-02

Title: A holistic approach to PET-degrading enzyme design: from QM/MM and MD simulations to ML and experiments

Speaker: Birgit Strodel, Heinrich Heine Universität Düsseldorf

Authors: Anna Jäckering, Tilman Hoffbauer; Birgit Strodel

In view of the increasing pollution from plastic waste, there is growing interest in developing an environmentally friendly enzymatic degradation process for plastics. The current bottlenecks are the limited activity and stability of enzymes for plastic waste recycling under industrial conditions. Our goal is to develop better variants of PET-degrading enzymes (called PETases). We started our approach by studying the binding behavior of PETases to PET and found that adsorption to the PET surface can be easily captured by classical molecular dynamics (MD) simulations, while the entry of a PET chain into the active site requires enhanced sampling due to the energy barriers involved. Therefore, we performed Hamiltonian Replica Exchange MD (HREMD) simulations that allowed us to sample hundreds of insertion events and thus identify the amino acids that facilitate or hinder PET entry into the binding site. Based on free energy profiles of amino acid-PET interactions, we generated mutations that were first tested in silico to assess their PET binding abilities using HREMD and catalytic efficiency using QM/MM calculations before characterizing the most promising mutation candidates in the wet lab of our collaborators at the University of Greifswald. In parallel, we developed an easy-to-use software tool called TransMEP that employs transfer learning by feature extraction with Gaussian process regression to predict the effect of mutations for protein engineering from existing data. We have demonstrated the power of TransMEP for other proteins and plan to apply it in the future to the dataset we are currently building for PET-degrading enzymes.

Abstract ID: IN-AS-02

Title: Accelerating the development of battery materials by high-performance computing

Speaker: Karsten Albe, Technische Universität Darmstadt

High-energy secondary batteries are currently spreading into a wide range of mobile and stationary applications. The potential for optimizing established liquid electrolyte-based lithium-ion batteries (LIB), however, is diminishing and thus alternative technologies need to be developed. In this context all-solid state batteries (SSB) relying on solid electrolytes show a lot of promise. Various material components are considered for SSB: Li metal anodes have high energy densities, but are difficult to process and suffer from contact losses, sulfur electrolytes exhibit excellent conductivies and are malleable, but exhibit a limited electrochemical stability, while Ni-rich layered oxides (NMC, NCA) remain the primary choice as cathode active material (CAM) containing, however, the critical elements cobalt. In order to better understand and eventually improve the properties and interplay of the various components of a SSB, computer simulations play a key role. In this contribution, I will highlight how the combination of ab-initio methods with classical and molecular dynamics as well as Monte-Carlo simulations allows to predict discharge curves, interface stabilities and ionic conductivities. A particular focus will be on developments, where the combination of high-performance computing with machine learning methods becomes a game changer.

Abstracts Contributed Talks

Abstract ID: CT-LS-01

Title: Curving molecules - on the fascinating role of lipids in COVID-19 vaccines and cell membranes

Speaker: Rainer Böckmann, Friedrich-Alexander-Universität Erlangen-Nürnberg

TBA

Abstract ID: CT-LS-02

Title: Cellular level modeling of membranes and proteins: large-scale application of mesoscopic simulations for organization and endocytosis

Speaker: Mohsen Sadeghi, Freie Universität Berlin

Biomembranes achieve their multitude of function in an organized and collaborative interplay with membrane-associated proteins. This organization is also present at the larger scale in the complex network of interacting proteins in e.g. the tegument of viral particles. Quantitative analysis of the dynamics of membranes interacting with a network of proteins in a consistent model that incorporates kinetics as well as protein structural information and flexibility is essential in fully describing these processes. Here, we present our dynamic framework for modeling membranes and proteins [1, 2], which includes our novel approach to hydrodynamic coupling [3]. We present results on protein aggregation [4, 5], spontaneous formation of membrane invaginations by membrane-bending toxins [6], and the first model of the human cytomegalovirus describing the organization of proteins in the viral tegument [7]. We make the case for how large-scale mesoscopic simulations achievable with high-performance computing platforms reach the biologically-relevant spatiotemporal scales.

[1] Sadeghi & Noé, Nat. Commun. (2020) 11:2951.
[2] Sadeghi, Weikl & Noé, J. Chem. Phys. (2018) 148:044901.
[3] Sadeghi & Noé, J. Chem. Phys. (2021) 155:114108.
[4] Sadeghi & Noé, J. Phys. Chem. Lett. (2021) 12:10497-10504.
[5] Sadeghi, Soft Matter (2022) 18:3917-3927.
[6] Sadeghi, bioRxiv (2022) 515891.
[7] Bogdanow et al Nat. Microbiol. (2023).

Abstract ID: CT-LS-03

Title: Multiparametric Quantitative Fluorescent Microscopy for the systems biology of endocytosis

Speaker: Yannis Kalaidzidis, Max Planck Institute for Molecular Cell Biology and Genetics Dresden

Authors: Yannis Kalaidzidis, Marino Zerial

Endocytosis serves multiple key functions in the cell, tissue development, and homeostasis. Our group focuses on understanding the molecular design principles that underlie the endocytic pathway and its contribution to intracellular signal transduction and interpretation. We employ a multidisciplinary approach to analyze this problem at different scales, including the molecular, cellular, and tissue level, using a combination of experimental and theoretical approaches. The aim of this project is to gain insights into the self-organization principles that govern tissue structure and function through a systematic functional analysis of cellular machineries across scales. We use cells in culture and mouse liver as a model to investigate endocytosis and aim to provide an integrated, multi-scale view ranging from the molecular to the tissue level.

Throughout our project, we have made significant advancements in quantitative image analysis algorithms. Particularly, we developed algorithms to investigate bile canaliculi (BC) formation both in laboratory settings, model living organisms and human tissue samples obtained from patients with primary sclerosing cholangitis (PSC), non-alcoholic steatohepatitis (NASH) and hepatocellular carcinoma, which were graciously provided by our collaborators at the university hospital of the TU Dresden. To gain a better understanding of liver disease progression, we expanded our computational analysis of BC to identify a common structural characteristic associated with the progression of liver diseases. Our findings suggest that structural alterations in the BC network lead to increased bile pressure, which in turn result in the reprogramming of hepatocytes and metabolic deregulations of liver tissue.

Abstract ID: CT-LS-04

Title: Learning an image with a homeostatic network using structural plasticity

Speaker: Marvin Kaster, Technische Universität Darmstadt

Author: Marvin Kaster

Co-authors: Fabian Czappa, Markus Butz-Ostendorf, Felix Wolf

Learning and memory are usually associated with Hebbian and synaptic plasticity. However, recent work showed that silent memory engrams can be formed using a homeostatic mechanism. We investigate how homeostatic models can learn more complex information and demonstrate this by learning a grayscale image.

We use a homeostatic model (Model of Structural Plasticity (MSP)) where each neuron seeks to maintain a steady fire rate. If a neuron is too much below or above the target activity, it grows/retracts synapses. We arrange excitatory neurons in a grid where a single neuron represents the gray value of a single pixel in the image. Additionally, we add randomly distributed inhibitory neurons. We stimulate a neuron with different intensities depending on the gray value of the associated pixel-resembling the persistence of a sensory stimulus in higher associative cortical networks in a more abstract sense- and observe that the outgrowth of synaptic elements depends on the intensity of the stimulation and, therefore, on the gray value of the pixel.

Nevertheless, the network is in an equilibrium maintaining the same fire rate for each neuron. The limited number of axons from inhibitory neurons results in an unequal distribution of inhibitory dendritic input of the excitatory neurons. Hence, the higher inhibitory input of the shorter stimulated neurons requires a higher amount of excitatory input to counter it and reach the same total synaptic input. Visualizing each neuron's number of dendritic elements enables us to reconstruct the original image with some noise.

Abstract ID: CT-LS-05

Title: Simulation-based Inference for Large-Scale Tissue Simulations

Speaker: Julian Herold, Karlsruhe Institute of Technology

Authors: Eric Behle, Julian Herold, Alexander Schug

While clinical imaging of tissues focuses on macroscopic tumors, many experiments investigate only small clusters of cells. We aim on providing a scale-bridging link by performing large scale tissue simulations. We employ highly parallelized code in an HPC setting to simulate mm-sized virtual tissues such as embryogenetic zebrafish tissue or breastcancer tumors with more than a million µm-resolved individual cells. We deploy Cells in Silico (CiS), which combines a cellular potts model with an agent based layer and is thus capable of accurately representing many physical and biological properties, such as individual cell shapes, cell division, cell motility, interactions with the extra-cellular matrix etc.

Using a model with such a strong representational capacity poses the task of adjusting a large number of parameters to reproduce experimental findings. Prior work has attempted to characterize the similarity between experimental and simulated data by extracting different features and using statistical tests to establish a distance measure. This work highlights how this difficult task can be circumvented by training neural networks to distinguish between experimental and simulated data while simultaneously optimizing the model parameters to maximize the error rate of the network.

Abstract ID: CT-LS-06

Title: HPC Facilitates Insights into the Genetic Complexity of Human and Bacterial

Speaker: Ali Al-Fatlawi, Technische Universität Dresden

Author: Ali Al-Fatlawi, Michael Schröder

In the era of genomics, understanding the intricate genetic underpinnings of complex diseases and drug resistance is paramount in life sciences. This talk highlights our use of High-Performance Computing (HPC) to advance comprehensive genomic data analysis. Through HPC, we have analyzed massive amounts of DNA and RNA sequences from human and bacterial sources, revealing critical mutations and gene expression patterns associated with various cancers and bacterial phenotypes in different environments. The journey from raw data to clinically relevant insights showcasing the power of HPC in health and environment applications. Furthermore, recent advancements in AI, particularly in protein folding prediction using AlphaFold, have further facilitated the applicability of training huge models with millions of parameters. At TU Dresden, we have successfully leveraged HPC ZIH to drive high-throughput heavy analysis, leading to innovations and opportunities for better research in life science. We are honoured to be invited to present a showcase of our HPC activities and recent publications in the field of Life Science.

Abstract ID: CT-LS-07

Title: Contact Maps in RNA Structure Prediction: Much more than pure Simulation Acceleratorssticity

Speaker: Christian Faber, Forschungszentrum Jülich

Author: Christian Faber

Co-authors: Alexander Schug, Oskar Taubert, Utkarsh Upadhyay, Benjamin Kotton

Predicting the spatial structure of non-coding RNA (ncRNA) is an important task for understanding fundamental processes in living nature. Physical force fields are used to infer the structure from a sequence using simulations on high-performance computers. However, the best results are obtained by incorporating evolutionary data via a binary mapping of contacts. The same phenomenon can be seen in protein structure prediction, where the groundbreaking AlphaFold2 also incorporates this step.

Much work has been done in the past to optimise the algorithms for simulations, but what are good contacts and why are these contacts important in the first place is an unsolved puzzle. To find answers, we tried different contact map topologies on a well-defined test set of ncRNAs. We also looked at using fewer, but wisely chosen contacts and how this can improve prediction. To obtain our results, we ran many simulations for comparison on the high performance cluster JUWELS with the RNA folding software SimRNA and used convolutional neural networks (CNN) to select contacts.

Our results suggest that it is important to pay more attention to the selection of contacts, especially when developing machine learning algorithms. Furthermore, good contacts not only ensure faster folding in the simulation, they are actually essential for correct folding. It seems that it is the additional constraints that bring the physical force field into the more correct form.

Abstract ID: CT-LS-08

Title: Large-Scale Synthetic Forest Point Cloud Generation on HPC Systems

Speaker: Ali Doosthosseini, Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen GWDG

Co-authors: Hauke Kirchner, Dorothea Sommer, Julian Kunkel

Forests play a vital role in sustaining life on earth. The study of morphology and distribution of trees within forests provides insight into their growth patterns and is essential for understanding forest evolution and threats. Various projects aim to image the forest at high-resolution and point clouds are frequently used to represent the three-dimensional structure and arrangement of trees. These large amounts of data typically require the use of HPC systems for processing.

Currently, acquisition of high-quality point cloud data is hindered by the cost of obtaining and operating imaging equipment such as drones and LiDAR sensors. Furthermore, the identification and segmentation of individual trees in point clouds present a challenge for feature extraction and subsequent analysis. To prepare for real-world high-resolution images, we developed SynForest, a workflow that generates high-quality point cloud datasets of synthetic forests. Our approach produces realistic patterns by combining forest growth simulation using ForestFactory and modeling trees based on real data. Additionally, the scanning method, trajectory, and sensor properties are simulated in Helios++ to approximate real imaging techniques. Through parallelism and task distribution on HPC systems, we are able to produce arbitrarily large datasets efficiently.

Our results demonstrate that synthetic datasets produced using SynForest closely resemble real-world data and offer a practical and cost-effective alternative for the study of large-scale forest environments and tree structures, with potential applications in deep learning.

Abstract ID: CT-LS-09

Title: Data-driven virus discovery by deep mining of raw sequencing data in high throughput

Speaker: Chris Lauber, Hannover Medical School

Authors: Chris Lauber, Stefan Seitz

The Sequence Read Archive (SRA) is a publicly accessible repository of unprocessed sequencing data hosted by the National Center for Biotechnology Information (NCBI). The SRA currently contains 14.5 million next generation sequencing (NGS) experiments that comprise more than 14 petabyte of data and represent the global sequencing effort of the scientific community. We have established high performance computing (HPC)-based workflows to efficiently and sensitively screen SRA data for the presence of known and novel viruses that may have been sequenced as a by-product of sequencing the host genome or transcriptome. Our data-driven virus discovery approach involves profile Hidden Markov Model (pHMM)-based sequence homology searches, targeted viral genome assembly and sequence-based taxonomic classification and is run on the HPC cluster Taurus of TU Dresden, allowing a high degree of parallelism. We have applied this approach to almost 1 million SRA datasets and present several examples of our findings, which include 35+ thousand RNA virus sequences most of which are novel, a divergent family of viruses that are the closest known relatives of hepatitis B viruses, the largest known RNA virus genome of 54 kilobases, coronaviruses with bisegmented genomes, relatives of the pandemic African swine fever virus, and human anelloviruses with potential disease associations. Our data-driven approach is currently transforming the field of virus discovery and is redefining our knowledge about viral diversity on the planet. The approach can be extended to other mobile genetic elements and to cellular organisms.

Abstract ID: CT-LS-10

Title: Ushering in a New Era for Data Processing in XNAT Using HPC

Speaker: James Philip Bowden, Department of Medical Informatics Universitätsmedizin Göttingen

Co-authors: Hendrik Nolte, Nicolai Spicher

Biomedical data produced by diagnostic techniques such as medical imaging and biosignal monitoring is extremely important in identifying and understanding many diseases. In the last decades, there has been steady progress in the clinical value of this data but also in required storage.

Hence, to be able to develop new methods to analyze these large datasets, proper data management is key. During the last 20 years, XNAT, the Extensible Neuroimaging Archiving Toolkit, has established itself as a standard choice as a research data management system to organize physiological datasets, even beyond imaging techniques.

Once the desired datasets are organized within the hierarchical structure of XNAT, the next step is typically to do some processing. Here, XNAT has a rich history, which started with its own built-in workflow engine. This was followed by the "Container Service", a mechanism to export data stored in XNAT into a Docker container, to execute arbitrary commands on the data. This service was continuously developed to support different backends, like Docker Swarm or Kubernetes. However, these setups reach their limits when drastically scaling out. In order to serve more demanding use cases, integration of HPC systems is required.

In this talk, we present a novel method to dispatch jobs from an XNAT instance to an HPC system. It was developed together with NHR@Göttingen and deployed on Emmy/Grete. As use cases, we give information on sleep stage classification from sleep laboratory data and on the prediction of neurological recovery based on biosignals acquired during coma.

Abstract ID: CT-LS-11

Title: Single-cell sequencing-based analysis of the tissue T cell microenvironment using high performance computing

Speaker: Kathrin Luise Braband, Institute for Immunology, University Medical Center Mainz

Authors: Kathrin Luise Braband, Michael Delacher, Andreas Henkel

In recent years, the development of high-throughput sequencing technologies has enabled the cost-efficient study of (epi)genome, transcriptome, interactome, and antigen receptor variation studies, amongst others. In our recently-funded transregional research center TRR355, several laboratories across our partner sites in Munich and Mainz as well as in affiliated institutes investigate aspects of Treg cells in different tissue microenvironments. Many of those projects rely on the characterization of lymphoid and non-lymphoid tissue Treg cells using state-of-the-art sequencing-based analytical methods such as single-cell chromatin accessibility mapping, single-cell gene expression analysis, or characterization of the T cell receptor expression landscape. In addition, novel amendments to existing technologies (e.g., barcoding of different samples, quantitative surface protein analysis using barcode-labeled antibodies) or entirely new technologies (e.g., physical cell-cell interaction, spatial transcriptomics and epigenomics) will be adapted. In addition to the optimization of wet-lab protocols, we will optimize bioinformatic analysis workflows for datasets generated using these protocols, which will include a centralized storage platform with computational capacity, a user-friendly interface for data processing, and a tool for the generation of output files for data visualization and further analysis. All bioinformatic processes are operated on the MOGON NHR Cluster in Mainz. The combination of protocol standardization and optimization with bioinformatics workflows and data visualization in our Treg commons platform will increase the comparability of datasets across laboratories to strengthen collaboration, reduce the overall experimental cost, analysis time and animal numbers, and provide the opportunity to compare datasets across labs and institutes.

Abstract ID: CT-LS-12

Title: Using SecureHPC to Process Sensitive MRI Health Data

Speaker: Philip Langer, Universitätsmedizin Göttingen

Author: Philip Langer

Co-author: Hendrik Nolte

Driven by the progress of data and compute-intensive methods in various scientific domains, there is an increasing demand from researchers working with highly sensitive data to have access to the necessary computational resources to be able to adapt those methods in their respective fields. To satisfy the computing needs of those researchers cost-effectively, it is an open quest to integrate reliable security measures on existing High Performance Computing (HPC) clusters. The fundamental problem with securely working with sensitive data is, that HPC systems are shared systems that are typically trimmed for the highest performance -- not for high security. Since new vulnerabilities are being continuously discovered, solely relying on the traditional Unix permissions is not secure enough.

In this talk, I present the usage of SecureHPC at University Medical Center Göttingen (UMG). Our use case consists of a large amount of high dimensional data, namely 3D and 4D images of the human brain to be analyzed via CPU/GPU calculations. The main difficulty in this approach is that magnetic resonance imaging (MRI) images of the head are inherently unanonymisable, as the brains shape is unique for every person, similar to a fingerprint, thus the data will remain highly sensitive in any case. For this reason we cannot use an external provider for the hardware intensive calculations needed for our project. GWDGs SecureHPC enables us to use high performance hardware while keeping the data within a secure enclosure, completely shut off from outside access, even in case of a vertical privilege escalation.

Abstract ID: CT-LS-13

Title: Coarse-Grained Simulation Model of Protein TDP43 Liquid-Liquid Phase Separation Behavior

Speaker: Xiaofei Ping, Johannes Gutenberg-Universität Mainz

Authors: Xiaofei Ping, Lukas Stelzl

TDP-43 (TAR DNA-binding protein 43) is a protein that involves in the pathogenesis of several neurodegenerative diseases, including Alzheimer's disease (AD), Amyotrophic lateral sclerosis (ALS), and Frontotemporal lobar degeneration (FTLD). The formation of protein aggregates in the brain and spinal cord of patients is underpinned by liquid-liquid phase separation (LLPS). Molecular dynamics simulations offer a powerful computational tool for investigating the behavior of molecules at the atomic or molecular scale, providing a detailed understanding of the driving force behind phase separation and aggregate formation. Two main types of simulations, atomistic and coarse-grained, are utilized in the study of molecular systems. Atomistic simulations provide a high level of detail and accuracy by representing each atom in the molecule. In contrast, coarsegrained models, such as the near-atomic resolution MARTINI and the residue-level Hydropathy scale (HPS) models, enable the investigation of larger systems and longer timescales. In the context of TDP-43 and neurodegenerative diseases, molecular dynamics simulations can provide insight into the phase separation behavior of TDP-43 condensates and provide avenues for understanding neurodegenerative diseases

Abstract ID: CT-AS-01

Title: From Petascale to Exascale: eigensolver ELPA library for atomistic simulations

Speaker: Hermann Lederer, Max Planck Computing and Data Facility Garching

Authors: Andreas Marek, Petr Karpov, Hermann Lederer

For many ab-initio atomistic simulations an efficient eigensolver for symmetric or hermitian matrices is a key element. In the last decade the ELPA library has evolved as highly efficient and highly scalable eigensolver package for all major HPC architectures and is in use by a large number of important ab-initio simulation packages. ELPA is "made in Germany" with the main developers from the Max Planck Society, the Technical University of Munich, and the University of Wuppertal. The ELPA development has been supported by two BMBF grants and currently plays a crucial role in the EU NOMAD Center of Excellence for the Exascale enabling of DFT simulation packages. In this talk the current progress of enabling the ELPA library for Exascale computing will be presented and extended scalability limits will be shown. Furthermore, the software design to support the different GPU architectures from the vendors AMD, Intel, and Nvidia will be discussed and we will show results of the first large-scale runs on the AMD GPU based top European Supercomputer Lumi. In this talk, high level features and newest developments for CPU and especially GPU based compute nodes are presented and we will discuss some preliminary results for employing GPU vendor specific communication libraries such as Nvidia's NCCL library together with a MPI implementation.

Abstract ID: CT-AS-02

Title: Towards efficient three-body calculations in molecular process engineering

Speaker: Philipp Neumann, Helmut-Schmidt-Universität Hamburg

Author: Philipp Neumann

Co-authors: Johann Duffek, Fabio Gratl, Samuel Newcome, Amartya Das Sharma, Markus Mühlhäußer, Alex Hocks

Typically, molecular dynamics simulations rely on pairwise particle interaction schemes. Depending on accuracy and computational requirements, efficient algorithms and parallelization schemes are available to implement these, including Linked Cells and Verlet lists for purely short-range interactions or fast multipole and Barnes-Hut methods for long-range interactions.

While pairwise interactions have shown to be a powerful model for a vast range of applications ranging from biology and life sciences to process engineering, they do have shortcomings, which, for example in molecular process engineering or thermodynamics, can result in inaccurate material property predictions.

One approach to address some of the shortcomings is to extend molecular simulation by non-additive three-body interactions, which, however, exhibit significantly higher computational loads than pairwise interaction schemes.

In this contribution, I lay out computational challenges with regard to three-body calculations and comment on our approach to incorporate three-body functionality into our molecular simulation software AutoPas/ls1 mardyn. AutoPas/ls1 mardyn is meant to simulate large systems of small, rigid molecules. Amongst others, the software supports automated algorithm selection (auto-tuning) at node-level and load balancing at distributed node-level. I will first explain some relevant features of AutoPas/ls1 mardyn, then describe the general three-body scenario and show first results on how to improve performance for these systems.

Abstract ID: CT-AS-03

Title: Atomistic simulations in thermodynamics

Speaker: Jadran Vrabec, Technische Universität Berlin

Atomistic simulations give access to the full range of thermodynamic properties of matter and allow for the study of processes with extreme temporal and spatial resolution. Equilibrium and non-equilibrium molecular simulations in the context of engineering science are presented. Coalescence of droplets and mass transport over the vapor-liquid interface due to evaporation are discussed as dynamic scenarios. Looking at a systematically selected series of fluids, the influence of molecular anisotropy and quadrupolar moment on their evaporation is studied by NEMD. For that purpose, ls1 mardyn Autopas is employed as a computational tool that has an outstanding scalability, being co-developed by our group. On thermodynamic properties, transport coefficients for diffusion are addressed, which are notoriously scarce in the literature because of experimental challenges. Employing the Green-Kubo formalism, the full range of transport coefficients can be obtained from equilibrium MD simulations. Looking at technically relevant mixtures, transport diffusion in the form of the Maxwell-Stefan and Fick approaches are studied. For that purpose, ms2 is employed as a computational tool that gives access to a large variety of time-dependent and time-independent properties, featuring MD and Monte Carlo simulations. Being parallelized on several levels, it is well-suited for HPC infrastructure so that an excellent time-to-solution can be achieved for typical investigations in thermodynamics. It is the result of a longstanding cooperation between process engineers and informaticists, maintained by our group.

Abstract ID: CT-AS-04

Title: Predicting Vibrational Spectra of Condensed Phase Systems

Speaker: Martin Brehm, Universität Paderborn

Vibrational spectroscopy is one of the most important tools of analytical chemistry. Predicting vibrational spectra via quantum chemistry is a standard technique since around 50 years, but mostly for isolated molecules or small clusters in vacuum, treating solvation only implicitly. However, it is well known that these spectra can be very sensitive to explicit solvent effects, e. g. if the solvent forms hydrogen bonds to the solute. In this contribution, I will discuss a more modern approach to vibrational spectroscopy, which is based on ab initio molecular dynamics (AIMD) simulations [1,2]. By performing such simulations of condensed phase systems in periodic boundary conditions, vibrational spectra of liquids can be reliably and accurately predicted in a black-box manner, including explicit solvent effects, realistic line shapes, and the most important anharmonic effects. The approach is currently applicable to predicting infrared, Raman, vibrational circular dichroism (VCD), Raman optical activity (ROA) [3], and resonande Raman spectra [4]. While the resulting spectra are often of high quality, the underlying simulations require large amounts of computer time, and therefore significantly profit from modern developments of highly scalable and efficient algorithms for quantum chemistry.

[1] M. Thomas, M. Brehm, R. Fligg, P. Vöhringer, B. Kirchner, Phys. Chem. Chem. Phys. 2013, 15,6608–6622.

[2] M. Brehm, M. Thomas, S. Gehrke, B. Kirchner, J. Chem. Phys. 2020, 152 (16), 164105.

[3] M. Brehm, M. Thomas, J. Phys. Chem. Lett. 2017, 8 (14), 3409–3414.

[4] M. Brehm, M. Thomas, J. Chem. Theory Comput. 2019, 15 (7), 3901–3905.

Abstract ID: CT-AS-05

Title: Differentiating GATE/Geant4 with Derivgrind

Speaker: Max Aehle, University of Kaiserslautern-Landau (RPTU)

Author: Max Aehle, Nicolas Gauger

In this talk, we present our work on differentiating Geant4, a Monte-Carlo simulation toolkit for the passage of particles through matter. We have applied our novel algorithmic differentiation (AD) tool. Derivgrind to the medical imaging application GATE (built on top of Geant4) to compute derivatives in a proton computed tomography setup. Specifically, the derivatives of hit coordinates of a proton in a detector with respect to the beam energy were evaluated, and successfully validated against numerical difference quotients.

Opposed to other AD tools, Derivgrind operates on the machine code of the target executable and libraries, by means of the heavyweight dynamic binary instrumentation framework Valgrind. One advantage of this approach is that Derivgrind requires less integration efforts, making it well-suited for these kind of exploratory studies of AD in new application domains, such as the fundamental sciences. But we also share a look on challenges of the machine-code-based approach related to code optimizations that improve performance, but "hide" real-arithmetic operations from AD tools.

Abstract ID: CT-AS-06

Title: Mechanistic insights into receptor mediated G protein activation

Speaker: Peter W. Hildebrand, Universität Leipzig

Authors: Hossein Batebi, Guillermo Pérez Haernández, Makaía M. Papasergi-Scott, Xiangyu Liu, Patrick Scheerer, Jesper Mosolff Mathiesen, Brian Kobilka, Georgios Skiniotis, Peter W. Hildebrand

Heterotrimeric G proteins are activated by G protein-coupled receptors (GPCRs) that mediate the exchange of guanine nucleotide in the Gα-subunit. We are investigating the mechanism of receptor-mediated G protein activation to obtain insights into the sequence of events that govern this important process. This process is divided into (i) functional association of a G protein with the receptor, (ii) nucleotide exchange, and (iii) G protein activation. To obtain an atomic-level description of the sequence of these events, we combine classical µ-second molecular dynamics simulations with time-resolved cryo-EM. Simulation of the association of the G protein with the β2-adrenoceptor describes the events that eventually lead to the ejection of GDP from the Galpha subunit, the rate-limiting step during G protein activation. Transmission of the signal to the G protein occurs via a novel receptor interface confirmed by site-directed mutagenesis and functional assays. From this β2AR-GsGDP intermediate, the G protein must undergo an in-plane rotation against the receptor to reach the β2AR-GsEMPTY state. Combining time-resolved cryo-EM with molecular dynamics simulations reveals that such rotation in the opposite direction occurs during the final steps of Gs protein activation. Thus, a corkscrew binding and unbinding pattern appears to underline the nucleotide exchange of G proteins by GPCRs. Our analysis sheds new light on the steps of receptor-mediated G protein activation and extends the limited view of nucleotide-free snapshots to include additional states and structural features responsible for signal transduction and specificity6 of G protein coupling.

Abstract ID: CT-AS-07

Title: Managing training data for ML Potentials

Speaker: Jan Kloppenburg, Aalto University

Authors: Jan Kloppenburg, Miguel Caro

Working on modern HPC infrastructure it is not always straight-forward to achieve acceptable performance for different problems. Small jobs (under whole-node occupancy) up to many-node jobs can be managed in an easy-to-use fashion almost right out of the box that is existing in many different python implementations. Computing a training data set for a ML potential and the application of it will serve as examples for small and large jobs with the number of computations being dozens of millions. Complex integration of user-defined workflows and job bundling are simple examples of how to construct abstract workflows for automated computation of target specific databases.

Abstract ID: CT-AS-08

Title: Physics-Informed Machine Learning for Addressing Challenges in Static and Time-Dependent Density Functional Theory

Speaker: Karan Shah, CASUS, Helmholtz-Zentrum Dresden-Rossendorf

Authors: Karan Shah, Attila Cangi

In this talk, we explore the potential of Physics-Informed Machine Learning (ML) in addressing key computational tasks in both static and time-dependent Density Functional Theory (DFT and TDDFT). The talk will focus on two projects that employ advanced ML techniques, specifically Physics-Informed Neural Networks (PINNs) and Fourier Neural Operators (FNOs), to tackle these complex tasks.

In the first part of the presentation, we examine the use of PINNs and FNOs in addressing the intricate density-to-potential inversion problem in static DFT. By employing these methods as alternatives to conventional inversion schemes, we demonstrate enhancements in predictive transferability and speed. We highlight the applications to exactly solvable models, such as soft-Coulomb systems, illustrating their potential as accurate and rapid data-driven surrogate models.

In the second part of the talk, we discuss the application of PINNs to accelerate TDDFT calculations. By incorporating the fundamental physical constraints of the Time-Dependent Kohn-Sham equations directly into the learning process, PINNs offer a unique way to fuse the power of deep learning with the nuances of TDDFT. We demonstrate the performance and generalisability of PINN solvers on the time evolution of model systems across varying system parameters, domains, and energy states.

By integrating physics and machine learning, these projects shed light on promising new directions for addressing challenges in DFT and TDDFT. The methods developed here have the potential to accelerate (TD)DFT workflows, enabling the simulation of large-scale calculations of electron dynamics in matter exposed to strong electromagnetic fields, high temperatures, and pressures.

Abstract ID: CT-AS-09

Title: Unraveling the atomic structure of silicon monoxide using high-performance computing and machine learning

Speaker: Linus Carl Erhard, Technische Universität Darmstadt

Author: Linus Carl Erhard

Co-authors: Jochen Rohrer, Karsten Albe, Volker Deringer

The structure of silicon monoxide has been discussed in the literature for about a century. It was found to be a mixture of silicon and silicon dioxide with nanometre-sized grains only twenty years ago. However, fully atomistic models that capture the nanostructure of silicon monoxide have not been available. Density functional theory (DFT) calculations are accurate enough to describe the interface between the grains, but size constraints prevent efficient calculations for the required grain sizes. Therefore, we use a multiscale approach that maps from quantum mechanical calculations to a fully atomistic model. High-performance computing allows us to perform more than ten thousand highly accurate DFT calculations, which we then use to train a machine-learning interatomic potential. Finally, this potential is used in molecular dynamics simulations to generate realistic structural models of silicon monoxide containing hundreds of thousands of atoms. These simulations take more than a week to run on several hundred cores.

Abstract ID: CT-AS-10

Title: Solid-solution modelling using first-principles methods: Case studies for (Cu,Ag)2ZnSnSe4 and (FA,Cs)PbI3

Speaker: Daniel Fritsch, Zuse Institute Berlin

Our quest to find new materials to be utilised in technological applications leads us more and more towards solid solutions between different materials. These solid solutions allow for the fine-tuning of desired material properties, but also pose additional problems in experimental characterisation and theoretical modelling. While we're able to deal with fractional occupancies of Wyckoff positions in experimental investigations, this is not the case for theoretical materials modelling based on density functional theory, and we have to resort to additional methods to properly model the structural, electronic, and optical properties of solid solutions [1].

Here, we're using first-principles calculations based on density functional theory to shed some light into the structure-property relations in the (Cu,Ag)2ZnSnSe4 [2] and (FA,Cs)PbI3 solid solutions (FA: formamidinium). While (Cu,Ag)2ZnSnSe4 only requires the mixing over different Wyckoff positions, in (FA,Cs)PbI3 we additionally have to account for the rotations of the FA cation. For both systems, in order to simulate the different concentrations within the solid solution, we're employing a supercell approach [3]. All our structure models are geometry optimised using the SCAN exchange and correlation functional, while electronic and optical properties have been obtained employing the more accurate hybrid functional HSE06 [4].

[1] D. Fritsch, in Crystallography in Materials Science, edited by S. Schorr and C. Weidenthaler, De Gruyter (2021).

[2] D. Fritsch and S. Schorr, J. Phys. Energy 3, 015002 (2021).

[3] D. Fritsch, Appl. Sci. 12, 2576 (2022).

[4] D. Fritsch, B. J. Morgan, and A. Walsh, Nanoscale Res. Lett. 12, 19 (2017).

Abstract ID: CT-AS-11

Title: Atomistic simulation based investigation of structure-property relations in silicon oxycarbides

Speaker: Niklas Leimeroth, Technische Universität Darmstadt

Co-authors: Jochen Rohrer, Karsten Albe

Silicon oxycarbides are of interest for a wide range of applications inlcuding energy storage, protective coatings and biomedicine. This unrivaled versatility stems from their highly tunable microstructure and composition. However, their structure and its relationship to thermodynamic and mechanical properties, is not well understood due to their complexity at the nanoscale. In this talk, we will present how molecular dynamics simulations based on modern machine learning potentials fitted to high-troughput ab-initio data can be used to reveal atomistic details of structures and how these relate to macroscopically measurable properties.

Abstract ID: CT-AS-12

Title: Towards understanding of exotic quantum phases from material realistic calculations

Speaker: Michael Winter, Universität Hamburg

Co-authors: Tim Wehling, Dominik Benner

Transition metal dichalcogenides (TMDs) are gaining significant interest due to their layered-material nature and the observation of exotic quantum phases, e.g., superconductivity and Mott physics, as well as the prediction of multi-knob tunability of these phases.

In this project, we work towards the general understanding of TMD-[hetero]bilayers from quantum-lattice models in twisted supercells. We carry out large scale and wide parameter range ab initio calculations including plane wave density functional theory, density functional perturbation theory, and subsequent electron-phonon interaction calculations. Upon an in-depth analysis of the resulting model, we want to address the question about possible origins of zero-resistance states realised in twisted bilayers of tungsten diselenide.

I will report on the current state of the model construction, addressing physics questions and computational pit-falls usually encountered along the way.

Abstract ID: CT-ABS-01

Title: Vahana.jl - A framework for large-scale agent-based models

Speaker: Steffen Fürst, Freie Universität Berlin

Vahana.jl is a new open-source HPC framework for large-scale agent-based models (ABM) of complex social systems, written in Julia. Vahana is based on a discrete dynamical systems formulation referred to as a synchronous graph dynamical system (SyGDS), which is a generalization of cellular automata (CA). Vahana extends the SyGDSs concept so that even complex models like the Mobility Transition Model (https://github.com/CoeGSS-Project/motmo) or MATSim Episim (https://github.com/matsim-org/matsim-episim-libs) can be expressed as a SyGDS.

(Discrete) spatial information can be added using Vahana functions that insert the grid cells as vertices in the graph and have access to a mapping from the Cartesian index of the underlaying space to the corresponding vertex. In the extended version, there are several types of vertices and edges, and each type can have its own state space. Edges not only determine the neighborhood, but can also be used as messages (if they are directed to another edge) or as memory (if they are a self-loop). There can also be multiple transition functions (which are the equivalent of a rule in a CA), and each transition function can act on a different subgraph.

Parallelization is done through the Message Passing Interface (via the MPI.jl package), but this is hidden to the user. Any model developed with Vahana can be automatically computed in parallel by simply starting the simulation via mpirun.

Abstract ID: CT-ABS-02

Title: Towards a digital twin to simulate bee health for Germany

Speaker: Jürgen Groeneveld, Helmholtz Zentrum für Umweltforschung Leipzig

Author: Jürgen Groeneveld

Co-authors: Jasmin Krebs, Tomas Martinovic, Tuomas Rossi, Anna Wendt, Volker Grimm

Honey bees (Apis mellifera) provide important services by producing honey and pollinating commercial crops but also a wide range of wild flowers. Honey bees face multiple stressors such as intensive agriculture, diseases and temporal limitation in nectar and pollen supply. Digital Twin applications can help to assess the quality of a given landscape for honey bee performance and honey production. In the BioDT project we aim to develop such a Digital Twin exploiting freely accessible land use data, weather data and the open source simulation model BEEHAVE. The computational challenge is to run the BEEHAVE model which is implemented in NetLogo on an HPC (LUMI) to be able to apply it for all of Germany. To do so the software required for executing the model (NetLogo, Java, R with nlrx and other packages) have been bundled in a Docker container image that can be pulled and executed on LUMI through Apptainer/Singularity and on a cloud through Docker. The main performance bottleneck of this use case is the large number of input data that requires processing. To overcome this, the execution of the containerised BEEHAVE model has been parallelised on the HPC (LUMI-C) over individual inputs by using HyperQueue. This way the use case can be executed in parallel on hundreds or thousands of cores, leveraging the large computing capacity of LUMI-C.

Abstract ID: CT-ABS-03

Title: Towards a digital twin to simulate bee health for Germany

Speaker: Björn Goldenbogen, Humboldt Universität zu Berlin

Co-authors: Edda Klipp

When COVID-19 began its global spread, the need for mathematical models to understand and predict disease dynamics and patterns became paramount. While compartmental models are effective for simulating strong outbreaks in highly connected populations, they may fall short when dealing with low infection numbers and heterogeneous connectivity, in particular during early stages of an outbreak.

In response, we developed GERDA, an open-source, geospatially referenced, demographic, agent-based model. GERDA captures the stochastic nature of disease spread within heterogeneous populations, where agent attributes, including age, influence infection probabilities. Our research, conducted using GERDA, revealed insights, notably in low-incidence scenarios. We found that non-pharmaceutical interventions can yield bimodal outcomes, introducing uncertainty into their effectiveness.

However, the computational demands of epidemiological ABMs are substantial. Simulating detailed synthetic populations exceeding 10,000 individuals is memory-intensive, and stochastic models require numerous repetitions for conclusive insights. To address this challenge, we explored various dimensionality reduction approaches. These included clustering highly connected individuals to reduce the number of interacting agents and temporal condensation, where we approximated infection probabilities over time. Our study demonstrates the potential of GERDA and highlights the tradeoffs and computational optimizations necessary when dealing with complex, fine-grained agent-based models in epidemiology.

Posters Abstracts

Abstract ID: P-AS-01

Title: DNA-Repair Mechanisms: Molecular Simulations and Computational Alchemy

Authors: Frank Beierlein, Senta Volkenandt, Petra Imhof

The DNA repair protein thymine DNA glycosylase (TDG) removes mispaired or damaged bases, such as oxidized methylcytosine, from DNA by cleavage of the glycosidic bond between the sugar and the target base flipped into the enzyme's active site. The enzyme is active against formyl-cytosine and carboxyl-cytosine, whereas the lower oxidized hydroxymethyl-cytosine and methyl-cytosine itself are not processed by the enzyme. To investigate the substrate specifity of TDG, we used extensive molecular dynamics simulations and thermodynamic integration of TDG complexed to DNA carrying one of four different (oxidized) methyl-cytosine bases methyl-cytosine (mC), hydroxymethyl-cytosine (hmC), formyl-cytosine (fC), or carboxyl-cytosine (caC), in extra- and intrahelical conformation, and in their amino- and imino-tautomeric forms. Our results indicate that discrimination of the oxidized methyl-cytosines does not take place in the initial complex formation before the base has been flipped out into the active site, and that imino-tautomers do not play a role in substrate recognition at this stage. For the extrahelical complexes, we observe a more favorable binding affinity of the higher oxidized forms, fC and caC, compared to the nonsubstrate bases hmC and mC. Despite rather comparable, reaction-competent conformations of the flipped bases in the active site of the enzyme, more and stronger interactions with active site residues account for the preferred binding of the higher oxidized bases. Overall, our computational results indicate that the enzyme discriminates the different oxidation forms of methyl-cytosine at the formation of the extrahelical complexes, and possibly also at a later chemical step.

Abstract ID: P-AS-02

Title: A domain-specific language for atomistic and molecular modeling

Authors: Ivan Kondov, Rodrigo Cortés Mejia

Co-authors: Marvin Müller, Nikolai Pfisterer, Sruthy Sreenivasan

Modeling and simulation on the atomic scale pose several challenges to their application on supercomputers. In particular, tackling the complexity of workflows, and enabling the reuse of code and data from the individual steps of these workflows are open issues. Adoption of workflow management systems (WMSs) has contributed significantly to overcome these challenges [1]. Nevertheless, domain scientists still need to invest great effort and considerable time in learning to use WMSs, high performance computing resources, and developing models based on domain-specific Python libraries. To increase the usability of the WMSs on supercomputers and the productivity of domain scientists, we have developed a use case-driven domain-specific language for atomistic and molecular modeling. This language and its supporting tools [2] facilitate the full life cycle of modeling, simulation, and data analysis including composing, parsing, interpreting, executing, and extending models thus allowing the reuse of models' data and methods. To maintain data persistence and provenance, the interpreter uses a back-end WMS equipped with a database. A transparent interface to Slurm enables executing computationally intensive steps on supercomputers. In addition, the tool set includes a language-specific Jupyter kernel enabling its use in Jupyter notebooks, for example by leveraging the JupyterHub user interface of HoreKa [3]. The language grammar has a modular structure that allows repurposing the language for use in other domains of computational science.

[1] J. Schaarschmidt et al., Advanced Energy Materials 2022, 12 (17), 2102638. https://doi.org/10.1002/aenm.202102638.

[2] VRE-Language, https://git.scc.kit.edu/virtmat-tools/vre-language

[3] NHR@KIT User Documentation, https://www.nhr.kit.edu/userdocs/jupyter

Abstract ID: P-AS-03

Title: One Ring to Rule Them All: Lugdunin's Disruptive Effects

Author: Marius F.W. Trollmann

Co-authors: Dominik Ruppelt, Claudia Steinem, Rainer A. Böckmann

Antimicrobial resistance represents a growing threat to global public health, underscoring the urgent need for novel strategies to counteract the spread of multi-resistant bacterial strains. Antimicrobial peptides (AMPs) have emerged as a promising alternative to common antibiotics for inhibiting bacterial growth without inducing new forms of resistance. Recently, the cyclic peptide lugdunin was isolated from nasal Staphylococcus lugdunensis and has shown a strong antimicrobial activity against several Gram-positive bacteria [1]. Lugdunin consists of six D,L-amino acids and a thiazolidine moiety. While maintaining membrane integrity, lugdunin was shown to enable proton translocation across the membrane [2]. However, the mechanistic mode of action of lugdunin on membranes is hardly understood.

Here, we applied atomistic molecular dynamics simulations to investigate lugdunin's differential interaction with a range of model membranes. Our results suggest that lugdunin easily penetrates the membrane interface region. We decipher the driving forces for lugdunin membrane embedment and in particular the role of the thiazolidine moiety, and discuss possibilities for channel formation. The presented simulations provide a basis for the future rational design of new macrocyclic thiazolidine peptide antibiotics with enhanced efficacy and safety profiles.

[1] Zipperer et al. 2016. Human commensals producing a novel antibiotic impair pathogen colonization. Nature. 535:511–516.

[2] Schilling et al. 2019. Synthetic Lugdunin Analogues Reveal Essential Structural Motifs for Antimicrobial Action and Proton Translocation Capability. Angew. Chem. Int. Ed Engl. 58:9234–9238.

Abstract ID: P-AS-04

Title: Unlocking the Secrets of UiO-MOFs: Exploring Porosity, Confinement Effects and Catalyst Performance

Authors: Sofia Kolin, Kristyna Pluhackova

Metal-organic frameworks (MOFs), consisting of metal-ion-based centers and organic linkers, have gained increasing attention in recent decades. The frameworks exhibit high porosity, tunable pore size and large surface area, which make them suitable for various applications including absorption, hydrogen storage, or catalysis. Among the well-known MOFs, the Zr-based UiO frameworks, consisting of a metallic center and 12 linkers of tunable length, stand out due to their exceptional thermal and pressure stability and solvent resistance.

Here, we constructed models of catalytically active UiO frameworks with linkers of varying lengths. In detail monophenyl, biphenyl or triphenyl linkers with/without the organic catalyst proline were utilized to generate frameworks with different porosity and catalyst loading of 25, 50, 75 or 100%. The systems are used to investigate adsorption in and diffusion through UiO-based MOFs using molecular dynamics (MD) simulations. Our goal is to unravel local diffusion and confinement effects of solvent, reactants and products and to predict maximal and optimal catalyst loadings for a specific combination of UiO-framework/catalyst/reaction kind.

Due to the large number of atoms and structural complexity of MOFs, coarse-grained (CG) MD simulations represent a meaningful simplification of the system by bundling atoms into so-called beads. Here, we parameterize the CG Martini3 force field for UiO-based MOFs. This process involves (i) establishment of the mapping strategy from AA to CG representation and (ii) fine-tuning the van der Waals properties of the CG beads through comparison with all-atom MD simulations. The CG parameters will boost our understanding and high-throughput characterization of MOFs.

Abstract ID: P-AS-05

Title: Incorporating electrostatic interactions in machine learning interatomic potentials

Authors: Alireza Ghasemi, Rahmatizad Khajehpasha Ehsan, Thomas D. Kühne

Machine learning interatomic potentials (MLIPs) have become popular in recent years and have been applied to a variety of materials. The standard methods have merely mathematical foundations driven by data without any sort of approximations to models in physics or chemistry. Furthermore, these methods can take into account short-range interactions because the descriptors, which encode atomic coordinates and compositions, depend only on neighbors within a localization region. There have been several efforts to incorporate physically motivated models into MLIPs among which the inclusion of electrostatic interactions has drawn much attention. This is due to the fact that electrostatic interactions are a ubiquitous term in reference methods based on quantum mechanics. We present a model built upon the charge equilibration process where the approximate reference density is obtained through an inverse scheme. The model takes into account only the long-range interaction and leaves the remainder of the interactions to a secondary machine learning process.

Abstract ID: P-AS-06

Title: Unveiling Biomolecular Proton Transfer and Tautomeric Equilibrium through High-Performance Computing

Author: Oscar Palomino Hernandez

Co-authors: Anna Riede, Lisa Johannknecht, Michel Lim, Paul Czodrowski

Proton transfer events are relevant in both biological and chemical processes. They play a crucial role in multiple processes such as biochemical reactions, pH regulation, acid-base reactions, catalysis, among others. Understanding proton transfer helps elucidate reaction mechanisms, develop new catalysts, and advance various fields in Biology, Medicine, and Chemistry.

In this work, we focus on the investigation of proton transfer events in the fields of biomolecular proton transfer and tautomeric equilibriums. Our research aims to shed light on the intricate mechanisms underlying these phenomena, leveraging advanced computational techniques and high-performance computing resources.

In the realm of biomolecular proton transfer, we have employed molecular dynamics simulations and quantum mechanics calculations to explore protonation and deprotonation events in biological systems. By elucidating the details of proton transfer pathways, we aim to unravel their fundamental roles in molecular recognition and coupled protonation-upon-binding from small molecules. Our presentation will highlight the computational methodologies utilized, including enhanced sampling techniques and constant-pH approaches, supported by HPC facilities.

Furthermore, we have delved into the field of tautomeric equilibrium, focusing on the dynamic interconversion between tautomeric forms in organic compounds. Through state-of-the-art quantum chemistry methods and machine learning approaches, we have explored the factors influencing tautomeric equilibrium and its implications in drug discovery. Our research demonstrates the power of high-performance computing in efficiently exploring large chemical spaces, accurately predicting tautomeric ratios, and uncovering structure-function relationships.

By presenting our findings at this conference, we look forwarde to foster interdisciplinary collaborations and engage in insightful discussions with fellow researchers.

Abstract ID: P-AS-07

Title: Structure Prediction of Iron Hydrides at High Pressure with Machine-Learned Interatomic Potentials

Authors: Hossein Tahmasbi, Kushal Ramakrishna, Mani Lokamani, Mandy Bethkenhagen, Attila Cangi

Understanding the composition of Earth's core and mantle is a major challenge in geoscience and materials science. The core is primarily made of iron, but its density is known to be slightly lower than pure iron. Hydrogen contributes to this density deficit, leading to significant interest in the properties and structure of iron hydrides under high pressure.

Previous studies have shown that the dhcp phase of FeH remains stable at lower pressures (10-40 GPa) but undergoes phase transitions to hcp and fcc phases at higher pressures. This study focuses on a theoretical exploration of the potential energy surfaces (PESs) of FeH under varying pressure conditions. The objective is to demonstrate the effectiveness of automated and systematic methods for training and validating transferable machine-learned interatomic potential (ML-IAP) using global optimization techniques. By utilizing this potential, which significantly reduces computational costs, the phase diagram of the stoichiometric Fe-H system is analyzed across a range of pressures.

To achieve this, we utilize the PyFLAME code to construct a highly transferable ML-IAP. With this accurate potential, the PESs of bulk FeH structures are systematically investigated through global sampling using the minima hopping method. This comprehensive exploration enables the prediction of stable and metastable iron hydrides from 0 to 100 GPa. Density functional theory calculations are conducted to refine the predicted structures and evaluate their dynamical stability. The findings of this study reveal a wide range of novel low-energy polymorphs of FeH at each pressure level, alongside the recovery of well-known structures in the literature.

Abstract ID: P-AS-08

Title: Chemical Accuracy at Low Computational Cost with σ-Functionals for the Kohn-Sham Correlation Energy

Authors: Andreas Görling, Egor Trushin, Jannis Erhard, Steffen Fauser

Recently, a new type of functionals for the Kohn-Sham (KS) correlation energy called σ-functionals that is based on the adiabatic-connection fluctuation-dissipation (ACFD) theorem was introduced. σ-Functionals are closely related to the well-known direct random phase approximation (dRPA). In the dRPA a coupling constant and a frequency integration of a function of the eigenvalues of the dynamic KS response function is carried out. In σ-functionals the dRPA expression in this integral is replaced by a function motivated by perturbation theory along the adiabatic connection [4]. In σ-functionals this function is optimized with respect to reference sets of reaction energies and other chemical properties in order to correct the errors resulting from neglecting the exchange-correlation kernel in dRPA.

σ-Functionals are applied in a post-self-consistent way using input orbitals and eigenvalues from a previous conventional DFT calculation. In Ref. [1] the PBE functional was used to generate the input quantities, in Refs. [2,3] also hybrid functionals (PBE0, B3LYP) were considered. The latter yielded mean absolute errors around or below 1 kcal/mol for reaction sets of reaction energies, barrier heights, and non-covalent interactions, and thus reach chemical accuracy.

Abstract ID: P-AS-09

Title: σ-functionals applied to molecules and solids

Authors: Andreas Görling, Joachim Paier

σ-functionals were recently introduced as accurate, yet computationally efficient alternatives to Kohn-Sham methods based on the direct random phase approximation (dRPA) to the correlation energy [1,2]. σ-functionals perform highly accurately when applied to main group compounds [1,2] and thus compete with high-level wavefunction methods. For instance, for reaction energies and energy barriers so-called chemical accuracy is reached, i.e., errors are within 1 kcal/mol. Here, we show computational results obtained using σ-functionals together with plane waves as basis set. Implementing σ-functionals requires dRPA, which is--among other periodic codes—available in the Vienna ab initio simulation package, VASP, and can be easily accomplished as "one-line add-on". We assess the plane-wave basis implementation by comparisons with GTO results for molecular systems. This is achieved by the calculation of corresponding atomization energies. Furthermore, applications to archetypal crystalline solids, especially metallic systems, have been carried out. With respect to the correlation energy contribution, plotting the associated integrand as a function of frequency, we found that required quadrature techniques must be applied carefully, i.e., involving denser meshes and integration to slightly higher energies, but this does not severely affect computational performance. Concerning solid state properties, it appears difficult to outperform dRPA when calculating lattice parameters and bulk moduli. However, with respect to binding or atomization energies of solids, σ-functionals improve the dRPA error bars by roughly 50%.

[1] E. Trushin, S. Thierbach, A. Görling, J. Chem. Phys. 154, 014104 (2021).

[2] C. Neiss, S. Fauser, and A Görling, J. Chem. Phys. 158, 044107 (2023).

Abstract ID: P-AS-10

Title: Density-functional calculations for two-dimensional metal-organic frameworks (2D-MOFs)

Author: Raviraj Chandreshbhai Mandalia

Co-authors: Andreas Görling

Self-assembled two-dimensional metal-organic frameworks (2D-MOF) offer unique properties paramount for heterogenous catalysis, gas sensing, and magnetic storage. The present work focuses on the computational investigation of the self-assembled metal-pyridyl coordination for free-standing 2D-MOFs and 2D-MOFs on different substrates. The molecule under study is 1,3,5-tris[4-(pyridin-4-yl)phenyl]benzene (ext-TPyB) with adatoms like Cu, Co, and Fe on graphene and on graphene put on different metal surfaces like Ir(111) and Ni(111). Literature reveals that the molecules self-assemble into two-fold and three-fold metal-pyridyl coordination. We perform first-principles calculations to study the underlying electronic structure of the systems. Potential energy surface for free-standing 2D-MOFs exhibiting two-fold and three-fold coordinated metal atoms are used to approximate the size of the system. The optimized geometries are used to determine magnetic properties using collinear density functional calculations. Such self-assembly of 2D-MOFs promises unique magnetic properties. The study provides preliminary information to set up new experimental investigations.

Abstract ID: P-AS-11

Title: Improving MD performance on HPC clusters through in-depth hardware knowledge and advanced program usage

Author: Anna Kahler

Co-authors: Thomas Zeiser, Gerhard Wellein

Most modern MD simulation programs run out of the box on HPC clusters and yield reasonable performance results. To shed some light on the backgrounds of optimized performance, we present three case studies from user support.

First, we present a reduction of hardware costs by 2/3: An REMD simulation with GROMACS reached 124 ns/day for 26 replicas on 12 dual-socket Intel Ice Lake. We were able to port this simulation to eight NVIDIA A40 GPUs while retaining a performance of 120.6 ns/day. Since the number of replicas is not a multiple of eight, porting required assignment of PP- and PME-tasks to the GPUs by hand.

The second case is about calling the GROMACS runtime correctly to obtain a performance gain, especially when dealing with a large simulation system of 2,600,000 atoms and a multiple GPU-setup. Starting from a performance of 11.8 ns/day on eight NVIDIA A40 GPUs, we nearly quadrupled performance to 20 ns/day on four A40 GPUs. Thus, about twice the performance on half of the resources by setting environment variables for improved GPU communication and adjusting runtime parameters.

Another proof of in-depth hardware knowledge is represented by our third case where ORCA underperformed on our high throughput cluster: A single numerical calculation of molecular frequencies took 76.4 hours to finish. Multiple setups on various CPU architectures followed by detailed examinations sped up this simulation to 11 hours on the same node; the statically linked OpenBLAS library falsely detected the underlying hardware.

Abstract ID: P-AS-12

Title: Sodium Channel Nav1.7: Molecular Insights Into Erythromelalgia Mutation

Author: Alessia Piergentili

Co-authors: Simone Albani, Anil Kumar Kalia, Giulia Rossetti, Angelika Lampert

Nav1.7 is a voltage gated sodium channel primarily found in the peripheral nervous system (PNS), playing a crucial role in detection of potentially painful stimuli. The gain-of-function mutation Q875E is associated with a neuropathic pain disorder known as Primary Erythromelalgia. It is known to cause hyperexcitability of the channel in sensory neurons by alleviating activation, slowing deactivation and enhancing response to small depolarizations of the membrane potential. Our previous experimental study suggested that this mutation stabilises the activated state of the voltage sensing domain I (VSD I) via the formation of a salt bridge between the gating charge R214 and the positive residue on the pore module E875. This pathophysiologic stabilization can be abolished by increasing the extracellular concentration of divalent cations (e.g. Ca2+ and Mg2+). We conducted High-Performance Computing based coarse-grained molecular dynamics (CG/MD) simulations on Nav1.7 with and without the Q875E mutation while exposing the channel to complex neuronal membrane composition and different concentrations of divalent ions. Our simulations convincingly support the hypothesis that Q875E indeed stabilizes VSD I in an active conformation by establishing electrostatic interactions between E875 and R214. We observed that an increase in the concentration of divalent ions can effectively reverse the effects of the pathogenic mutant due to a generic 'masking' effect of divalent ion solvation around opposite charge residues exposed to the solvent. This study sheds light on the molecular mechanisms underlying the pathophysiology of the Q875E mutation in Nav1.7, offering potential insights into the development of therapies targeting such pain disorder.

Abstract ID: P-AS-13

Title: Joint forces: Hypothesis generation for transmembrane protein oligomerization by combining AlphaFold2 and Molecular Dynamics simulations

Author: David Rosenberger

Co-author: Cecilia Clementi

The transport of ions through lipid bilayers via pores formed by oligomers of proteins often plays a critical role in the life cycle of a virus. Blocking the so-called viroporins have been successful drug targets for curing viral infections. The Dengue virus (DV) is a virus for which we currently cannot identify potential drug targets as it is unclear if e.g. similar viroporins are formed. AlphaFold2 (AF2) has emerged as a powerful tool to predict protein structures from sequence, but its performance on predicting transmembrane protein complexes remains unclear despite recent advancements. Therefore, we perform molecular dynamics (MD) simulations to predict the oligomeric complex of the transmembrane protein of DV (M-DV) in order to identify a potential novel drug target. We study the self-assembly of multiple individual chains starting from AF2 predictions of the monomer given the lack of a experimentally resolved structure. We utilize coarse-grained MD simulations to reduce the computational cost of the self-assembly simulations. Nevertheless, atomistic details are still needed to study the specific pore structure. This leads to the following computational protocol to shine a light if M-DV forms a viroporin, and if the viroporin has a biological function: 1.) Making a structure prediction for M-DV using AF2, 2.) reduce atomistic details via coarse-graining, 3.) perform self-assembly simulations for each possible complex (dimer, trimer, tetramer, pentamer), 4.) identify stable pore structure in each oligomeric complex and re-introduce atomistic details to study the pore structure in detail. Using this protocol hints towards a trimer formation of M-DV.

Abstract ID: P-AS-14

Title: Computational Investigations of MnPX3 (X: S, Se) Monolayers for Electrocatalytic Water Splitting

Authors: Jiajun Dai, Elena Voloshina, Beate Paulus

Low-dimensional materials like graphene or transition metal dicalcogenides have high surface to bulk ratios and therefore can provide a large number of active sites for catalysis. In this project we analyse the active sites for water splitting of the less investigated MnPX3 (X:S,Se) layered material.

A DFT approach at PBE+U+D2 level was used to study the structural and electronic properties of pristine and defective MnPX3 monolayers as well as their activity towards water and hydrogen evolution reaction (HER) catalytic performance. On the pristine structure, H2O is physisorbed whereas it prefers to adsorb stronger on at defect site and the dissociation process energetically favorable. Following the Norskov approach, HER catalytic performance is evaluated by calculating hydrogen adsorption free energy. Our calculation results demonstrate that defective 2D MnPX3 with low coordinated P shows significantly better HER performance compared to their pristine counterpart.

Abstract ID: P-AS-15

Title: Machine Learned Molecular Coarse-Graining of Protein Folding and Aggregation

Authors: Andrea Guljas, Nicholas Charron, Yaoyi Chen, Félix Musil, Andreas Krämer, Frank Noé, and Cecilia Clementi

Molecular coarse-graining (CG) is an invaluable tool for understanding protein dynamics, particularly in the study of large biomolecular systems. By leveraging state-of-the-art machine learning approaches, it is possible to develop a CG potential that effectively captures the multi-bodied effects that arise when atomistic degrees of freedom are integrated out. Such a potential can subsequently be used to model the molecular motions associated with proteinfolding and aggregation. To ensure that the model is transferable across either sequence or replica space, we fine-tune both the composition of the input data used for training, which consist of atomistic simulations, and the physicsinformed correction terms that are applied to the model. The resulting potential can be used in simulations to explore the dynamics of large systems across long timescales, and thus better understand the molecular properties associated with protein interactions.

Abstract ID: P-AS-16

Title: Enhancing Data Efficiency in Coarse-Grained Force Field Parameterization with Machine Learning

Author: Yaoyi Chen

Coarse-grained models extend the tractable time- and length-scale of molecular dynamics, widening its application in biomedical research. To accurately parameterize force fields for a reduced resolution, machine learning approaches based on variational force matching have been proposed. However, noisy coarse-grained forces put demands on a large number of conformational samples, especially for long polypeptides. We present two solutions to this data efficiency challenge. First, by leveraging the flexibility of force mapping operator, we can use a statistically optimized way to aggregate of all-atom forces. Alternatively, we employ normalizing flows, a generative deep learning method, to directly learn from the conformational distribution without relying on force information, commonly not recorded in simulation studies. Our successful applications on fast-folding proteins demonstrate the potential of these methods in achieving accurate and versatile coarse-grained force fields.

Abstract ID: P-LS-01

Title: Novel Antiviral Strategies: Structure-Based Design and Optimization of PG16-Antibody Derived CDRH3 Peptides against HIV1

Authors: Anselm H. C. Horn, Manuel Deubler, Lukas Weißenborn, Simon Leukel, Simon Schäfer, Jutta Eichler, Heinrich Sticht

Broadly neutralizing antibodies (bnAb) binding to viral fusion proteins offer promising strategies for protection from viral infections. Such antibodies can be used for passive immunization and are currently tested in clinical trials, but they are expensive and difficult to produce. As a more economic alternative, antibody-derived peptides may be used for this purpose.

Suitable antibody sequences were identified using a computational pipeline that identified 2050 interfaces of HIV-1 antibody-antigen complexes and revealed several promising candidate peptides, which were investigated by molecular dynamics (MD) simulations. The first peptides investigated by this MD-based optimization approach are from the sulfo-tyrosine containing bnAb PG16. Optimization of the peptide length is based on an energetic analysis of the complex interface, while the effect of peptide cyclisation was assessed via microsecond MD-simulations of the free peptides. This approach resulted in a high-affinity peptide ligand that exhibited a nanomolar affinity for the HIV-1 gp120 protein in laboratory experiments.

Abstract ID: P-LS-02

Title: Machine Learning Guided RNA Contact Prediction

Author: Utkarsh Upadhyay

Co-authors: Oskar Taubert, Christian Faber, Alexander Schug

For around 50 years, the primary focus of genomic research has been the development of efficient and accurate methods to predict the structure of proteins, which led to the birth of better sequencing techniques and databases. About 98% of the human genome (RNA, DNA) during this action was overlooked.

Many consider RNA merely as a messenger between DNA and ribosomes for making proteins. However, In the past few years, studies have revealed the existence of many non-coding RNAs which catalyse various biological processes; to gain detailed insights into these roles, we require the appropriate structure of RNAs. Recent years have led to breakthroughs in protein structure prediction via Deep Learning. The scarcity of RNA structures, however, makes a direct transfer of these methods impossible. Here, we present machine-learning techniques that can work with limited training data. We predict contact maps as a proxy to understand and predict RNA structure, they provide a minimal representation of the structure. We have worked on methods that took accuracy from 47%(DCA) to 77%(CoCoNet) and now to 87%(Barnacle) i.e. doubling accuracy while reducing false positives by five-fold. Further, research is going on to create much more efficient neural networks which make use of statistical physics and ML techniques like Attention mechanisms and Transformers. We are confident that this remarkable progress will reduce the sequence-structure gap for RNA.

Abstract ID: P-LS-03

Title: Learning Active Matter Behavior of Poxvirus Spreading through Time-lapse Image Generation by Denoising Diffusion Probabilistic Model

Authors: Gabriel della Maggiora, Artur Yakimovich

Complex dynamic processes occurring in nature may be captured by time-lapse imaging. However, understanding and reproducing these processes remains a challenge. These processes range from mass transfer in fluids to the complex behaviour of live active matter dynamics in cell motility driven by poxvirus infection spread in a monolayer of cells. Understanding these processes can be attempted through time-lapse sequence synthesis by means of generative modelling. Here, we present a novel method to predict behaviour from video sequences, where the underlying mechanics are governed by differential equations with known and unknown characteristics. Our method is an extension of residual video diffusion in which we learn an approximation of the underlying differential equation separating the drift term and the stochastic term. We evaluate it with the reaction-diffusion equation in which we hide the inhibitor variable from the model and the incompressible Navier-Stokes equation with a stochastic forcing parameter. Our model accurately predicts the inhibitor variable in the reaction-diffusion equations and the stochastic forcing parameter in the Navier-Stokes equation. Additionally, we evaluate the model's capability to learn distinct biological behaviours of the active matter when trained on time-lapse microscopy of poxvirus spread phenotypes. The results confirm the model's potential in capturing meaningful equation embeddings, thus contributing to a deeper understanding of biological dynamics. To assess the accuracy of poxvirus prediction, we measured the mean absolute error when close to the initial condition. To evaluate the generation of longer sequences, we employed a qualitative analysis in which our model achieved excellent results.

Abstract ID: P-LS-04

Title: Complex formation between Polyethylenimine and mRNA

Author: Jonas Lehnen

Co-authors: Friederike Schmid, Giovanni Settanni

Messenger RNA vaccines have proven invaluable in the fight against the COVID-19 pandemic. Among the vehicles for non-viral gene delivery Polyethylenimine (PEI) has attracted attention due to its high transfection efficiency. PEI binds to negatively charged mRNA forming polyplexes. These are nanoparticles (NP) of different sizes, depending on the pH used for their assembly as well as salt, PEI and RNA concentration. Small NPs have been shown to be critical for high transfection efficiency. We use coarse-grained molecular dynamics simulations to examine the effects of the various factors determining polyplex size and gain a better understanding of the processes involved in their formation, with a special interest on the effects of PEI concentrations way above the amount necessary to neutralize the mRNA, following up on recent experimental results.

Experimental and coarse grained simulation results are compared and results about the driving mechanisms responsible for controlling the size of NPs are shown. Of particular interest here is a change in the driving mechanisms of the NP size above and below the isoelectric point of the system.

Abstract ID: P-LS-05

Title: Spike Sorting using Dask: From Laptop to HPC

Author: Friedrich Schwarz

Co-authors: Andreas Neef, Jonas Franz, Fred Wolf

Spike sorting – the process of extracting neural activity of individual nerve cells from raw extracellular recordings – is fundamental in analysing neurophysiology data. Current algorithms often fail to cope with increasing data volumes and time constraints requiring solutions despite GPU utilisation.

Using Dask - an open-source Python library for parallel computing - we developed a flexible spike sorting algorithm effortlessly scalable from laboratory laptops to HPC systems. This approach combines several advantages, including resource allocation commensurate with data volume, reduced developing times in environments with fast-changing analysis requirements through local testing, and changeless subsequent HPC deployment. All this without losing the option of further GPU integration and accelerations. Moreover, it allows experimental groups to avoid purchasing expensive hardware systems by utilising central HPC facilities. Still, to approach real-time analysis, low latency data transfers from the measurement site to the HPC's scratch are needed. This is realised by choosing Zarr as storage backend instead of HDF5. I/O bounds can be further reduced by combining Dask's and Zarr's parallelisation capabilities, reaching greater general I/O throughputs than HDF5-based implementations for our data. If needed, a fsspec- and kerchunk-based implementation assures the seamless use of legacy HDF5 files.

The modular Python-based implementation assures a low-threshold integration of custom analyses without knowing the computational backend allowing a greater scientific audience to benefit from HPC performance.

The implementation's very competitive analysis runtime and stability permit extensive time-appropriate hyper-parameter tuning, e.g., with Bayesian-based AutoML methods, ensuring the best possible parameter constellation for varying experimental recording conditions.

Abstract ID: P-LS-06

Title: Identification and structural characterization of peptidic ligands for novel antiviral strategies against SARS-CoV-2

Author: Olena Denysenko

Co-authors: Manuel Deubler, Simon Schäfer, Anselm H.C. Horn, Heinrich Sticht

Broadly neutralizing antibodies that block viral entry are efficient in preventing infection, but they are difficult to handle. As a substitute, antibody-derived peptides may be used for this purpose. Peptides are smaller in size and therefore more likely more capable of reaching sterically shielded epitopes, which makes them good candidates for drug design. In our project, we aim to identifypeptides with antiviral activity by analyzing complexes between antibodies and SARS CoV-2 spike protein to identify energetic hot-spots of the interaction. For this energetic analysis we developed a specific pipeline to identify key interaction sites between SARS-CoV-2 and antibody structures, which provide the basis for peptide design.

We have already used this pipeline to characterize 826 experimental structures of antibodies in complex with the spike protein of SARS Co V-2. The energetically most favorable peptides are currently further characterized by molecular dynamics simulations. In addition, these novel peptides will also be used for creating bispecific ligands by covalently linking two peptide ligands that recognize neighboring sites of the antigen. From such an approach we expect an enhanced binding affinity to the spike protein of SARS Co V-2.

Abstract ID: P-LS-07

Title: Role of UTX in the regulation of hematopoietic stem cell fate decisions during cell cycle progression

Author: Anupam Sinha

Gene expression among other things is regulated by chromatin modification. Tri-methylation of H3K27 leads to down-regulation of gene expression. In this study we wish to understand the the dynamics of Haematopoietic stem cells (HSCs) heterogeneity with deletion for the gene Kdm6a , which functions as a histone demethylase for H3K27 and alters the expression pattern of multiple genes simultaneously. We have generated single-cell(sc) RNA-seq data from the Kdm6a deletion mutants and normal HSCs. After clustering of the data and cell-cycle scoring of the cell states we have mapped cell-cycle marker genes to the different clusters. We noticed an increase in the G2/M fraction among cell clusters in the deletion mutants as indicated by the strong expression of genes like Ccnb1. We have also mapped immune cell-lineage type gene expression signatures on to these clusters to delineate the biases in the developmental dynamics of normal and mutant cells towards lymphoid or myeloid lineages. The next steps will be to create dynamic models of the cell-state transitions by integrating the single-cell expression data with Boolean models. We would also like to study the multicellular dynamics of cell-clusters using agent-based modelling. In the HSC niche, our mutant cells interact with normal cells thereby influencing each others' behaviours. An important outcome of this study would be to determine how these interactions shape the dynamics of lineage development. Essentially, we wish to scale-up from genotype (deletion mutation) to dynamic pathway models to whole cell simulation, ultimately leading to the population based behaviour of HSC niche.

Abstract ID: P-LS-08

Title: Bottom-up machine learning models can compete with the UNRES model

Authors: Aldo S. Pasos-Trejo, Nick E. Charron, Felix E. Musil, Frank Noe and Cecilia Clementi

Coarse-grained (CG) models allow us to simulate biomolecular systems over timescales large enough to observe biologically relevant processes. When investigating protein dynamics, the UNRES model has exhibited success in both protein folding and structure prediction while having a simple functional form and using a low dimensional representation of its systems. A big drawback, however, is that different interactions of its force field have to be parametrized with different methodologies.

Recently, machine learning (ML) approaches have allowed us to build bottom-up CG models that have an accuracy that rivals the latest versions of the UNRES model. The bottom-up parametrization enables a more straightforward process for building and testing CG models while the less restrictive functional form of ML can make significant improvements in various scenarios when comparing with UNRES.

Abstract ID: P-LS-09

Title: Providing mechanistic insight into hemeproteins through quantum mechanical calculations of hemes in the protein environment

Authors: Jesse Jones, Maria Andrea Mroginski, Tobias Gensch, Iris Guo, Konrad Leich, Elizaveta Zhartovska

Heme proteins are one of the structural classes of proteins with some of the most influential and diverse functions of all non-amino-acid cofactors in proteins. Preliminary results have shown that the environment has a great effect on the physical properties of the heme, from which the mechanistic functions arise. Here we present calculations on heme cofactors at DFT level, which show promising results as to simple structural factors, the porphyrin torsions, greatly influencing the redox potential of the hemes. The redox potential is assumed to be the defining variable for most of the hemes reaction behaviours. The underlying torsion values show high predictive power for the redox potential, and could be used in turn to predict the reaction behaviour of the hemes.

Abstract ID: P-ABS-01

Title: Efficient Coupling of Highly Parallel Computational FLuid Dynamics Simulations with Machine Learning on Heterogeneous Architectures

Authors: Fabian Orland, Ludovico Nista, Tim Jeremy Patrick Karpowski, Christian Terboven, Heinz Pitsch, Federica Ferraro, Christian Hasse

Driven by the convergence of HPC and AI, deep learning (DL) models become an increasingly important tool to complement traditional numerical simulations across different domain sciences. Coupling of traditional HPC codes and DL models is challenging as heterogeneous hardware architectures consisting of CPUs and GPUs, or other specialized AI accelerators, need to be exploited efficiently. In this work, we present two use cases of deploying DL models into HPC simulation codes from the field of simulating turbulent reactive flows.

First, detailed chemistry calculations lead to solving stiff ODE systems that are computationally very expensive to solve. A cheaper alternative is tabulated chemistry using flamelet-generated manifolds. This however, imposes severe memory limitations though. A promising alternative is an artificial neural network that learns the non-linear relationship between thermochemical control variables and chemical source terms.

Second, the large eddy simulation formalism introduces a low-pass filter resulting in unclosed non-linear advection terms in the governing Navier-Stokes equations. To close these terms subgrid-scale models are required, for which DL models are investigated as alternatives to algebraic models.

To facilitate the deployment of such models into existing HPC simulation codes on modern, heterogeneous computer architectures, we developed a software library, called AIxeleratorService. This library provides an API abstracting from concrete machine learning framework APIs and hiding data communication between CPUs and GPUs to accelerate the DL model inference in a heterogeneous job.

We demonstrate the applicability of our library by implementing it into an existing OpenFOAM solver as well as an in-house solver called CIAO..

Abstract ID: P-ABS-02

Title: Comparison of the Numerical Flow Iteration to particle-based approaches for the Vlasov equation

Authors: Rostislav-Paul Wilhelm, Matthias Kirchhart, Manuel Torrilhon

Modelling of high temperature plasmas, as they arise in, e.g., fusion reactors or particle accelerators, often involves the high-dimensional Vlasov equation arising from kinetic theory. The curse of dimensionality as well as the weak-collisionality of the equation, which introduces filamentation into the solution, lead to long compute times and a large memory-footprint for state-of-the-art approaches as Particle-In-Cell (PIC) or Semi-Lagrangian (SL) schemes.

We present a novel approach, the Numerical Flow Iteration (NuFI), which uses an iterative-in-time reconstruction of the characteristics of the Vlasov equation. Instead of storing the distribution function, we propose to only store the lower-dimensional potentials of the electro-magnetic fields and evaluate the distribution function on-the-fly.

Due to its several orders smaller memory-requirement and high flop/Byte-rates the algorithm can run on moderate hardware, however, it also scales well on large compute clusters as the scheme is embarrassingly parallel and communication between compute-nodes can be kept to a minimum. Furthermore the scheme is norm-, entropy- and energy-conserving. To show-case above properties we compare NuFI to other particle-based approaches for several benchmark cases.

Abstract ID: P-O-01

Title: Efficient Graph Neural Network Integration for Cost-Effective Language Model Training in Graph-to-Text Generation

Authors: Färber Michael, Shuzhou Yuan

Large language models (LLMs) have proven highly effective in various downstream natural language processing (NLP) tasks. However, their extensive training and fine-tuning costs demand significant computational resources. To address this challenge, our research explores the integration of graph neural networks (GNNs) into LLMs, aiming to mitigate the training cost of LLMs while maintaining their performance.

In particular, we focus on the graph-to-text generation task and introduce GNNs as adapters within the attention layer of LLMs. During the training process, we freeze the parameters of the LLMs and solely update the parameters of the GNNs. By doing so, the model gains the ability to effectively capture the structural information present in the input data, enabling the generation of fluent and descriptive text. This novel approach allows us to achieve considerable savings in training costs, as we only update a smaller subset of parameters within the model, without compromising its overall effectiveness.

The proposed method holds significant promise and can be applied to a wide range of NLP tasks that involve structured data, such as social networks and syntax-based text. By adopting GNN-based adapters in LLMs, researchers and practitioners can tackle resource-intensive NLP tasks more efficiently, making large-scale language models more accessible and sustainable for a broader range of applications.

Abstract ID: P-O-02

Title: Implementation Techniques for SPMD Kernels on CPUs

Authors: Joachim Mathias Meyer, Aksel Alpay, Holger Fröning, Sebastian Hack, Vincent Heuveline

More and more frameworks and simulations are developed using heterogeneous programming models such as OpenCL, SYCL, CUDA, or HIP. A significant hurdle to mapping these models to CPUs in a performance-portable manner is that implementing work-group barriers for such kernels requires providing forward-progress guarantees so that all work-items can reach the barrier. This work provides guidance for implementations of single-program multiple-data (SPMD) programming models, such as OpenCL, SYCL, CUDA, or HIP, on non-SPMD devices, such as CPUs. We discuss the trade-offs of multiple approaches to handling work-group-level barriers. Thereby, we review a general design flaw in deep loop fission approaches, as used in the popular Portable Computing Language (PoCL) project, that makes them miscompile certain kernels.

For our evaluation, we integrate PoCL's "loopvec" kernel compiler into hipSYCL and implement continuation-based synchronization (CBS) in the same. We compare both against hipSYCL's library-only fiber implementation using diverse hardware: we use recent AMD Rome and Intel Icelake server CPUs but also two Arm server CPUs, namely Fujitsu's A64FX and Marvell's ThunderX2.

We show that compiler-based approaches outperform library-only implementations by up to multiple orders of magnitude. Further, we adapt our CBS implementation into PoCL and compare it against its loopvec approach in both, PoCL and hipSYCL. We find that our implementation of CBS, while being more general than PoCL's approach, gives comparable performance in PoCL and even surpasses it in hipSYCL. Therefore we recommend its use in general.

Abstract ID: P-O-03

Title: Automatic Differentiation of OpenMP with OpDiLib

Authors: Johannes Blühdorn, Nicolas Gauger

Automatic differentiation (AD) is an established tool in scientific computing for the efficient retrieval of accurate derivatives. The reverse mode of AD, also known as backpropagation in machine learning, enables, for example, sensitivity analysis, parameter identification, and discrete adjoint procedures for simulation codes from all application domains. To handle large-scale problems, these codes are typically run on high performance computing clusters and written according to parallel paradigms such as MPI or OpenMP, which require dedicated handling in terms of AD. We present OpDiLib, an add-on for operator overloading AD tools that enables the reverse mode differentiation of OpenMP parallel codes. It relies on OpenMP's tools interface (OMPT) to detect and augment OpenMP constructs in a fully automatic fashion. OpDiLib deduces a corresponding reverse parallelism and embeds it into the tape of the underlying operator overloading AD tool. For the sake of generality and applicability, OpDiLib addresses the inherent data races of shared-memory parallel reverse mode AD by defaulting to atomic updates on adjoint variables. To improve the performance, this can be lifted according to user knowledge in a fine-grained manner. Our benchmarks showcase the achievable speedup and scalability and demonstrate the impact of optimizations such as the elimination of atomic updates on adjoint variables, both in OpenMP parallel and in MPI-OpenMP hybrid parallel applications.

Abstract ID: P-O-04

Title: Supporting containerized HPC environments

Authors: Peter Faber, Aadesh Baskar

Container technologies have standardized and simplified the packaging of software applications along with their prerequisites. The container-based approach simplifies application deployment, makes version management easier, and upgrades quicker. Along with these benefits, the usage of container-based approaches in HPC also increases the flexibility to deploy the applications in the cloud and serverless architectures. We configured a prototype HPC cluster based on the container-based Warewulf framework to run containerized MPI applications and used it to configure and evaluate different setups and workflows, in particular running Quantum Computing simulations using the Intel Quantum Simulator. The areas of focus are the creation of container images, configuration of container engines, user management, file storage, scheduling jobs, container orchestration, and provisioning of operating systems. We aim at providing a first map of possibilities for implementing a purely container-based HPC environment.

To that end, we evaluated technologies like Podman, Warewulf, SLURM, and Kubernetes.

We configured a cluster to let unprivileged users start container-defined MPI applications and proceeded to use Warewulf to provision stateless worker nodes with the root filesystem taken from a given container image. We are currently building a cluster combining both the above-mentioned approaches: A base operating system provisioned by Warewulf on stateless nodes capable of running containerized MPI applications scheduled by SLURM or orchestrated by Kubernetes.

Our presentation shows the different setups and workflows we configured along with their effects on flexibility and ease of deploying the application stack on an HPC cluster.

Abstract ID: P-O-06

Title: Improving Job Efficiency by Gamification

Authors: Andreas Henkel, Severin Hasselbach, Jens Rutten

In the context of High Performance Computing (HPC), where a large amount of users work on and share the same hardware cluster, monitoring is a core, if not essential, component. By collecting information about hardware and software performance, status, efficiency, workload, duration of jobs etc., possibilities for further improvement are identified while troubleshooting is simplified as well. To provide insights into cluster status, workload and performance as well as job-specific performance and job statistics, we introduce MOgoN SUbmit and moNitoring (MONSUN), a web-based interface built with Flask and Elasticsearch, which supplies a clear and simple, yet powerful frontend for users and administrators. For users, MONSUN will offer optimization of cluster usage by means of gamification. Average job efficiencies will be ranked per user and simple hints based on elementary heuristics will be offered. By this, the users are encouraged to improve their position in the job efficiency ranking by following the hints on revising their jobs and, as a byproduct, learn to use the cluster more efficiently and improve the cluster utilization as a whole.

Abstract ID: P-O-07

Title: Novel Comparison of Noise Models Simulation with Runs on Quantum Computer

Authors: Mahmoud Abuzayed, Helena Liebelt, Rui Li

Noise significantly impacts the performance of quantum computers, making it imperative to consider its effects in discussions about practical quantum computing. Benchmarking approaches cannot succeed without addressing noise and its impact on measurements. Comparing noise models with real quantum computer runs helps validate simulations and understand hardware limitations.

Accurately simulating a quantum processor requires realistic noise models, but identifying the appropriate model and performing the simulation computationally is challenging. This proposed work is part of a collaborative research effort involving the Deggendorf Institute of Technology, Leibniz Supercomputer Center (LRZ), Atos, and IQM. Firstly, we will deploy a quantum circuit with the noise model of the IQM superconducting 5-qubit system on the Quantum Learning Machine (QLM) emulator, then on the real quantum hardware. The unique advantage of both the QLM and the IQM 5-qubit system being located at LRZ premises is the opportunity to compare algorithm simulations on the QLM with the newly developed noise model to runs on real hardware using the same algorithms.

The primary aim of this research is to utilize the quantum emulator (QLM) for benchmarking real hardware at LRZ. The comparison will involve characterizing and modeling the noise processes in the hardware to improve the models for realistic simulation of digital quantum circuits. This study focuses on accurately modeling quantum noise and assessing the robustness of quantum operations. By doing so, the research aims to understand the impact of the noise model on the solution, identify differences from the ideal solution, and analyze the introduced errors.