

## NHR- Graduate School: Research Fields

### List of Contents

|  |           |
|--|-----------|
| <b>NHR4CES@RWTH – Aachen / NHR4CES@TUDa – Darmstadt.....</b>   | <b>3</b>  |
| 1#NHR4CES: Data Management and Data Science and Machine Learning (RWTH/TUDa) .....                                 | 3         |
| 2#NHR4CES: Parallelism and Performance (RWTH/TUDa).....  | 3         |
| 3#NHR4CES: Visualization (RWTH) .....  | 4         |
| 4#NHR4CES: Energy Conversion (RWTH/TUDa) .....   | 4         |
| 5#NHR4CES: Fluids (RWTH/TUDa).....   | 4         |
| 6#NHR4CES: Materials Design (RWTH/TUDa).....   | 4         |
| 7#NHR4CES: Digital Patient (RWTH) .....  | 4         |
| <b>NHR@FAU – Erlangen.....</b>   | <b>6</b>  |
| 1#FAU: Scalable Solvers .....  | 6         |
| 2#FAU: Performance Engineering from Single Core to Highly Parallel Architectures.....                              | 6         |
| 3#FAU: Molecular Simulations of Bio-Molecular Complexes.....   | 6         |
| 4#FAU: Development of new electronic structure methods combining high accuracy with computational efficiency ..... | 7         |
| 5#FAU: High-throughput Simulations for a Mechanistic Understanding of Crystallization Processes.....               | 7         |
| <b>GWDG – NHR@GÖTTINGEN – Göttingen .....</b>  | <b>8</b>  |
| 1#Göttingen: HPC-Cloud-Convergence .....   | 8         |
| 2#Göttingen: Efficient Data-Driven Workflows .....   | 8         |
| 3#Göttingen: Data-Center Twin .....  | 8         |
| 4#Göttingen: Scalable Big Data Analytics .....   | 8         |
| <b>NHR@KIT – Karlsruhe .....</b>   | <b>9</b>  |
| 1#KIT: High performance computations for multiphysics models .....   | 9         |
| 2#KIT: Combined numerical and parallel methods.....  | 9         |
| 3#KIT: Biologically-inspired artificial intelligence algorithms .....  | 9         |
| 4#KIT: Sustainable Software Development .....  | 10        |
| 5#KIT: Data-intensive computing .....  | 10        |
| <b>PC2 – Paderborn Center for Parallel Computing – Paderborn .....</b>   | <b>11</b> |
| 1#PC2: Sparse Linear Algebra with FPGAs.....   | 11        |
| 2#PC2: Scalable and Performance-Portable Scientific FPGA Applications.....   | 11        |
| 3#PC2: Methods for Atomistic Simulations on Heterogeneous Computer Architectures.....                              | 12        |
| 4#PC2: Methods for compressing smooth volumetric data on grids .....   | 12        |
| 5#PC2: Electronic Structure Theory on Quantum Computers.....   | 13        |

|  |           |
|--|-----------|
| <b>ZIH – NHR@TUD – Dresden</b> .....   | <b>14</b> |
| 1#TUD: <i>Software Tools for Parallel Performance Analysis and Optimization</i> .....  | 14        |
| 2#TUD: <i>Task Parallel Programming Models</i> .....                                   | 14        |
| 3#TUD: <i>Data-Driven HPC Workflows</i> .....  | 14        |
| 4#TUD: <i>Machine Learning</i> .....   | 15        |
| 5#TUD: <i>Energy Efficient Computing</i> .....   | 15        |
| 6#TUD: <i>Machine Learning for Research Data Management</i> .....                      | 15        |
| <b>NHR@ZIB – Berlin</b> .....  | <b>16</b> |
| 1#ZIB: <i>Hybrid Quantum-Classical Algorithms for Combinatorial Optimization</i> ..... | 16        |
| 2#ZIB: <i>Large-scale Explainable AI</i> .....   | 16        |
| 3#ZIB: <i>Semantic I/O with Object Stores in HPC</i> .....                             | 16        |
| 4#ZIB: <i>Modelling Excitations in Novel Materials</i> .....                           | 17        |
| 5#ZIB: <i>Cross-Platform and AI-enhanced Molecular Dynamics</i> .....                  | 17        |
| 6#ZIB: <i>Portable and Energy-Efficient Computing on Data-Parallel Platforms</i> ..... | 17        |
| 7#ZIB: <i>Massive Simulation of Agent-Based Models</i> .....                           | 18        |

# NHR4CES@RWTH – Aachen und NHR4CES@TUDa – Darmstadt

In NHR4CES, RWTH Aachen University and Technical University Darmstadt join forces to combine their strengths in HPC applications, algorithms and methods, and the efficient use of HPC hardware. Our goal is to create an ecosystem combining best practices of HPC and research data management to address questions that are of central importance for technical developments in economy and society. We support the development of scalable algorithms and software for the investigation, development, design, construction, evaluation, and production of engineering applications with a particular focus on engineering and materials science as well as engineering-oriented physics, chemistry and medicine.

## **1#NHR4CES: Data Management and Data Science and Machine Learning (RWTH/TUDa)**

Data-intensive applications are getting more and more important in the context of HPC infrastructure. First, classical HPC users who in the past used simulation-based and thus computation-intensive methods rely more and more on data-driven methods such as ML and AI. Second, as another trend, we see that also the user base of HPC infrastructure is growing since many disciplines that did not use HPC in the past, plan to use the HPC infrastructure for tackling data-intensive problems using the scalable resources. As such, the goal of the CSGs Data Management and Data Science and Machine Learning (RWTH/TUDa) is to develop novel methods to enable classical HPC users but also new user groups to implement data-intensive applications on top of the scalable HPC infrastructure.

For this, the aim of the CSGs is to develop a range of different methods that can broadly be classified into two main directions: (1) The existing data services such as long-term data storage are today completely decoupled from the HPC infrastructure. As such a first goal is to better integrate the existing data services into the available HPC infrastructure (RWTH). (2) As a second goal, we aim to develop novel methods to enable users without data science skills to build data-intensive applications by automating data engineering tasks such as data transformation, logging and cleaning as well as automating the ML model building e.g. via neural architecture search and making ML and process mining approaches differentiable (TUDa, RWTH). By automating data engineering and ML model building, we enable classical HPC users as well as new user groups who are no experts in data science and ML to build data-driven applications much faster and with much lower (manual) overhead than today.

## **2#NHR4CES: Parallelism and Performance (RWTH/TUDa)**

Performance and Parallelism-related research is being executed at RWTH and TUDa, in several cases in joint projects. The groups invent methods and algorithms to assess and improve the correctness, performance and parallelism of HPC applications and develop tools (MUST, Archer, Extra-P, Disco-POP) to put theory into practice. Based on work with real-world applications, they propose new approaches to program modern HPC architectures, bring these into standardization bodies such as MPI and OpenMP, and extend the corresponding runtime systems. In doing so, they regularly explore and characterize new system and microprocessor architectures.

### **3#NHR4CES: Visualization (RWTH)**

Visualization allows for the efficient exploration and analysis of data in virtually all domains. The rapid development of hardware and simulation software, however, introduces a variety of challenges for scientific visualization. This includes finding new approaches for the visualization of complex and uncertain data types, dealing with large data, or analyzing running simulations with in-situ methods. The goal is to provide researchers with efficient tools to gain better insights into their data and techniques may range from providing immersive environments to defining new data features.

### **4#NHR4CES: Energy Conversion (RWTH/TUDa)**

We develop HPC-ready reactive CFD (rCFD) software and methods by a co-design process. Such rCFD software is key for exploring physics and efficient virtual designs of energy conversion devices as part of the energy transition, a global challenge with major economic and social impact. Highly optimized numerical approaches are developed, tested, validated, and packaged in different forms for dissemination. Resulting specialized HPC-modules are highly parallelized and optimized for the targeted Tier II architectures, also providing an efficient usage on Tier I machines. Software quality is ensured together with the CSGs by Continuous Integration (CI). Two simulation frameworks will be used, the DNS code Ciao developed at RWTH Aachen University (Prof. Pitsch) and the open-source code OpenFOAM, customized for LES and 3D URANS applications at TUDa (Prof. Hasse).

### **5#NHR4CES: Fluids (RWTH/TUDa)**

The detailed analysis of flow problems is essential in energy conversion, environmental sciences, climate research etc. Besides experiments numerical investigations are to be performed to solve the scientific challenges which are related to multi-scale multi-physics problems such as particle-laden flows, gas-liquid flows, flows of complex and particulate matter, aerodynamics, aeroacoustics, combustion. To capture the details of such problems high-performance computing is a must. Furthermore, model reduction for multi-query applications, and data- and physics-based model integration are emerging topics. An entire spectrum of scalable discretization methods – finite volumes (Prof. Wolfgang Schröder, Prof. Jeanette Hussong, Prof. Suad Jakirlic), finite elements (Prof. Marek Behr), Lattice-Boltzmann (Prof. Martin Oberlack) – is considered.

### **6#NHR4CES: Materials Design (RWTH/TUDa)**

Molecular dynamics simulations, using high-performance computing architectures, allow to explore materials properties and processes with high spatial and temporal resolution. While running multi-million particle simulations nowadays can be done on a routine basis, the analysis of all trajectories is still a challenging matter. With the advent of novel data analysis methods based on machine learning algorithms, new and improved analysis methods are at the horizon. In our work, we address the development and implementation of analysis methods as well as the integration into community-based software tools. Possible projects include the simulation and structure recognition of grain boundary phases, glasses and complex microstructures.

### **7#NHR4CES: Digital Patient (RWTH)**

Efficient computational methods for development of digital patients representing disease progression, therapeutic options and efficiency of personalized therapies are essential for future digitalization in medicine, both in research and clinical health care. Digital patients require an optimal adaption of model structures to the medical needs, availability of

mechanistic knowledge and data. Moreover, digital patients exhibit multi-scale structures ranging from the molecular up to multi-organ level. They can be realized by hybrid integration of physics-based models, machine learning and molecular simulations. Molecular simulations as well as the identification of digital patients from data requires efficient adaption of HPC methods, e.g., Markov-Chain Monte Carlo simulations, to the specific requirements of hybrid ML-mechanistic models. These methods need to be adapted to the specific heterogeneous structures of digital patients and data.

# NHR@FAU – Erlangen

Research activities at the Erlangen National High-Performance Computing Center (NHR@FAU) focus on performance engineering, scalable solvers and atomistic structure simulations in chemistry, life science, materials science and physics. NHR@FAU is part of the Friedrich-Alexander-Universität Erlangen-Nürnberg.

## **1#FAU: Scalable Solvers**

In many HPC applications, solving sparse linear systems is the fundamental performance bottleneck. Only few algorithmic principles, such as, e.g., multigrid, are available to construct scalable solvers. Further acceleration is possible with techniques that avoid storing the system matrix. Such matrix-free techniques are one focus of research at NHR@FAU. These methods require that the solver is co-designed with the discretization method. For more applications to profit, further extensions are necessary, such as the surrogate matrix technique. Here the coefficients of the system matrix are approximated as a set of polynomials that can be evaluated efficiently on-the-fly. A central component in building highly efficient and easily usable implementations in this co-design process are modern code-generation (meta-programming) techniques which form another research focus of NHR@FAU. (Prof. Ulrich Rüde)

## **2#FAU: Performance Engineering from Single Core to Highly Parallel Architectures**

Performance Engineering (PE) is a structured, performance-model-based process for the structured optimization and parallelization of basic operations, algorithms and application codes for modern compute architectures. Performance models and performance tools are the core components of the PE process. The group uses and further develops the execution cache memory (ECM) model, which is a generalization of the well-known Roofline model and allows predictions of single-core performance and scaling within a multi-core chip. The interaction of node-level bottlenecks with highly parallel code execution is another recent area of research. Furthermore, simple-to-use performance analysis and modeling tools such as LIKWID or OSACA are developed at NHR@FAU. Hardware-efficient building blocks for sparse linear algebra and stencil solvers are strong topical foci of the PE activities of the center. (Prof. Gerhard Wellein)

## **3#FAU: Molecular Simulations of Bio-Molecular Complexes**

Bio-molecular complexes such as protein-protein, protein-DNA, or (other) protein ligand complexes are flexible entities whose conformational dynamics is important for their function. By molecular dynamics simulations, using high-performance computing architectures, these dynamics can be explored in atomic detail. However, depending on the process to be investigated, advanced methods such as enhanced sampling techniques or approaches with hybrid quantum mechanical and classical (QM/MM) potentials are needed, often also in combination. We are using a catalogue of molecular simulation techniques and data analysis methods to study bio-molecular complexes involved in important processes such as DNA repair, enzymatic reaction pathways, ligand-induced signaling, or immune processes at the plasma membrane. (Prof. Petra Imhof, Prof. Heinrich Sticht, Prof. Rainer Böckmann)

#### **4#FAU: Development of new electronic structure methods combining high accuracy with computational efficiency**

The workhorse for electronic structure calculations in chemistry, physics, and materials science are methods based on density-functional theory (DFT). Despite their enormous success in practical application, presently available DFT methods suffer from a number of shortcomings. Their accuracy does not reliably reach what is called chemical accuracy, i.e., an accuracy of 1 kcal/mol in reaction energies, and non-covalent interaction cannot be treated. A promising new type of DFT methods relies on the adiabatic-connection fluctuation-dissipation (ACFD) theorem. ACFD methods exhibit an unprecedented accuracy and applicability. These methods need to be implemented in highly efficient codes making full use of the opportunities high-performance computing offers. (Prof. Andreas Görling)

#### **5#FAU: High-throughput Simulations for a Mechanistic Understanding of Crystallization Processes**

The formation of materials structure can be rationalized, to first approximation, from geometric considerations. But while crystal structure is generally well understood, dynamic aspects of crystallization are much less accessible. The prevailing strategy in computational materials science is to simulate with ever higher accuracy. Here, we propose to investigate crystallization processes systematically across parameter spaces and aim at a mechanistic understanding that correlates the geometry of matter (structure) and interaction (chemistry) to details of ordering pathways at full atomic resolution. This goal is accomplished by establishing a crystallographic structure analysis and simulation infrastructure that rivals experimental synthesis and structure solution from scattering data. Analysis of trajectory data addresses long-standing open questions in the context of multi-step nucleation and directional growth. Applications are phase diagrams of intermetallics, industrial crystal growth, and fractional crystallization in geology. (Prof. Michael Engel)

# GWDG – NHR@GÖTTINGEN – Göttingen

The Gesellschaft für wissenschaftliche Datenverarbeitung Göttingen (GWDG) is a joint computing center and research facility of the Max Planck Society and the University of Göttingen. It provides infrastructures for data management and analysis, collaboration and computing for various national projects. This national high-performance data center with the Emmy supercomputer is a tier 2 center and offers advice and support to a diverse portfolio of users and is itself involved in research in application-related projects. The NHR center NHR@Göttingen is part of the National High Performance Computing (NHR) and a member of the NHR network.

## **1#Göttingen: HPC-Cloud-Convergence**

In recent years, classic HPC users have seen an ever-increasing interest in the public cloud that is used as part of traditional HPC workflows. There are many reasons for this, e.g. the usage of special hardware components in the cloud, or that data products created in a data center are shared to a scientific community - that continues to process it in the cloud using High-Performance Data Analytics. At the same time, a technology initially developed for the cloud is increasingly used in HPC centers and vice versa. Hence, cloud and HPC systems are converging. As part of this topic, synergistic effects in compute, storage, and management are researched. (Prof. Julian Kunkel)

## **2#Göttingen: Efficient Data-Driven Workflows**

The efficient, convenient, and robust data management and execution of datadriven workflows are key for productivity in computer-aided research. Still, the storage stack is based on low-level I/O that requires complex manual tuning. In this topic, we are researching means for lifting the abstraction to a new level paving the road for intelligent storage systems. One expected benefit of these systems is the prospect exploitation of heterogeneous storage and compute infrastructures by allowing to schedule user workloads efficiently across a system topology – a concept called Liquid Computing. Ultimately, intelligent compute and storage solutions can improve data handling for HPC but also Big Data Analytics workflows over time without user intervention and lead towards an era with smart system infrastructure. (Prof. Julian Kunkel)

## **3#Göttingen: Data-Center Twin**

Data centers contain various infrastructure to, e.g., steer airflow and cooling of housed systems. Understanding and optimizing the infrastructure but also the whole life cycle of IT components is important to improve efficiency and make computing greener. The modelling and simulation of the technical processes and relevant characteristics allows investigating the impact of various tuning options and enables what-if analysis to optimize planning in a given data center but also in future data centers. (Prof. Julian Kunkel)

## **4#Göttingen: Scalable Big Data Analytics**

Machine learning methods are omnipresent in our daily lives and are also increasingly used in scientific domains such as life science and the digital humanities. However, in the data center, HPC workflows could still benefit more from opportunities by machine learning and deep learning methods. As part of this topic, we research 1) how scalable data analytics workflows can be established for relevant scientific domains, and 2) how existing workflows can be augmented using deep learning, machine learning, and High-Performance Data Analytics, in general. (Prof. Julian Kunkel)



# NHR@KIT – Karlsruhe

Karlsruhe Institute of Technology (KIT) is a university of the state of Baden-Württemberg as well as a research center in the Helmholtz Association and creates synergies from serving both purposes. The NHR center NHR@KIT addresses four user communities from earth system science, materials science, engineering for energy and mobility research, and particle physics. Core HPC method competences are data-intensive computing, numerical algorithms, and sustainable software development. In HPC operations, NHR@KIT focuses on secure authentication and authorization.

## **1#KIT: High performance computations for multiphysics models**

High-performance computing is crucial for the numerical solution of multiphysics models, as they are used in diverse fields of application, and enables a treatment of the relevant processes in representative volume elements. To describe coupled processes such as heat and mass transport in materials, flow and heat transport or thermo-chemo-mechanical interaction, complex partial differential equations must be solved numerically. The result promises a prediction of property correlations that enables processes to be redesigned in a resource- and energy-efficient way through computer simulations. This requires a consistent parallelisation of the solution methods using MPI and OpenMP to implement fast parallel software solutions on high-performance computers. The goal is to capture processes on different length and time scales and to include correlations as homogenised parameters in a multi-step and multi-scale workflow.

## **2#KIT: Combined numerical and parallel methods**

Efficient use of high-performance computers for solving coupled dynamic evolution processes requires a combination of efficient numerical solution methods and parallel computation methods. The central application is the implementation of combined implementations for multi-scale simulations, in which processes take place on different time and length scales that must be resolved with different accuracy. The use of adaptive methods and multigrid methods is of central importance in this area. The aim of the research work is to parallelise these numerical methods via MPI and OpenMP and to implement efficient scaling simulations on HPC systems and to create a systematic hierarchical approach and a uniform simulation framework for multiscale processes that includes programming interfaces for data processing and numerical approaches in combination with a consistently parallel solution with high scaling properties.

## **3#KIT: Biologically-inspired artificial intelligence algorithms**

Biologically-inspired computing aims at solving complex computational problems by mimicking biological process or structures. Especially in complex optimization problems, as can be found in a number of artificial intelligence and machine learning applications, bio-inspired algorithms are recognized as state-of-the-art approaches. Examples for these algorithms include genetic and particle swarm optimization, spiking neural networks or feedback-free learning. NHR@KIT is looking for projects targeting, but not limited to, biologically-inspired neural architecture search, hyperparameter optimization, learning algorithms for (deep) neural networks and neuromorphic computing. A strong emphasis is put on the scalability of the researched biologically-inspired algorithms to large-scale computational applications, exposing inherent parallelism.

#### **4#KIT: Sustainable Software Development**

Most representative community codes and software infrastructures consist of large components developed over long periods by many different software developers. Many of these codes focus on a single type of computational resource, e.g., large clusters with standard CPUs and a very fast interconnection network. As the number of developers increases and a project takes longer to develop, the likelihood that bugs will be introduced into the code base increases because it is not possible to manually test all possible configurations and validate the results for each change. NHR@KIT is looking for projects that make scientific codes fit for the future, to make optimal use of the available computing resources, including porting, testing and benchmarking on new architectures.

#### **5#KIT: Data-intensive computing**

Enormous amounts of data are generated by ever more highly scaling simulations on HPC systems, but also by large-scale experiments. In order to analyse this data and gain new scientific insights from it, services for processing and evaluating such data are required, which go hand in hand with a corresponding infrastructure. The creation of an appropriate environment for the management and analysis of research data in view of the rapid growth rates of real and simulated data in conjunction with new data science methods such as machine learning also requires easy access to computing resources. The goal of NHR@KIT is to gain a deep understanding of the information content of experimental and simulated data and to build data-driven models to transform data into knowledge.

## PC2 – Paderborn Center for Parallel Computing – Paderborn

The Paderborn Center for Parallel Computing (PC2) is a central research institute and HPC center at Paderborn University with the mission to conduct interdisciplinary research at the intersection of computational science, computer science, and innovative computer architecture. Within the NHR Alliance, PC2 focuses on applications in computational physics and chemistry; numerical methods and libraries for heterogeneous computing systems; and energy-efficient computer architectures such as FPGAs and GPUs. The supervision of the proposed doctoral topics with a strong connection to natural sciences is carried out jointly with corresponding chairs from the Faculty of Natural Sciences.

### **1#PC2: Sparse Linear Algebra with FPGAs**

For addressing large-scale computational science and engineering problems, the use of sparse matrix/vector data representations is essential for keeping the memory and computational demand of the simulations tractable. The efficient execution of linear algebra operations on sparse data structures is however challenging because the predominant current computer architectures are optimized for dense rather than sparse linear algebra and also suffer from a widening gap between computational and memory throughput. This is problematic for memory-bound sparse operations. Therefore, CPUs or GPUs typically achieve only a small fraction of their peak performance for sparse linear algebra.

FPGAs can close this gap by customizing the internal memory architecture, tailoring the sparse matrix storage format, and chaining algebraic operations to avoid repeated transfers from/to external DRAM memory. First studies have demonstrated the potential of FPGAs for sparse operations for selected operations but so far there is no cohesive design method or library of composable matrix operations for FPGAs.

Examples for PhD projects in this area are: development of composable building blocks for sparse linear algebra; developing novel techniques for integrating the building blocks into more complex algorithms (e.g. solvers); study of arithmetic customization (low/mixed-precision computing); integration into widely- used open-source application codes.

### **2#PC2: Scalable and Performance-Portable Scientific FPGA Applications**

The use of FPGAs as accelerators has so far mostly focused on strong-scaling applications for single nodes. The next step for the adoption of FPGAs in HPC is to build scalable, parallel, multi-node applications for scientifically relevant application domains. For highly scalable applications, suitable network architectures and communication libraries are required that are usable directly from FPGA applications without routing all communication through the host. Also, runtime systems and middleware for parallel FPGA systems are still a wide-open research field.

Examples for PhD projects in this area are: development of communication infrastructures and libraries that are suitable for direct use from FPGA applications; Another open challenge is the performance portability of FPGA applications among different FPGA types or to other architectures such as CPUs and GPUs. While OpenCL and SYCL-based tools have brought great progress in terms of functional portability, performance optimization techniques still differ widely. In the context of scalable applications with FPGAs directly communicating via common standards, performance portability can also be important heterogeneous execution modes where different application kernels are flexibly mapped to the most suitable device.

Further examples for PhD projects in this area are: improvement of performance portability through code generation, or domain-specific languages; exploration of configurable and composable architecture-optimized libraries or template-based architecture adaptation; techniques for heterogeneous execution modes combining streaming and dataflow execution modes with block based or bulk synchronous execution modes.

### **3#PC2: Methods for Atomistic Simulations on Heterogeneous Computer Architectures**

The simulation of atomic interactions at different levels of physical accuracy and numerical demand ranging from force-field molecular dynamics over static and dynamic density functional theory up to quantum chemistry are a cornerstone for computational science in physics, chemistry, material sciences, and life sciences. Because of the extremely widespread use and computational demand of these methods, improvements in efficiency and performance in the underlying methods, algorithms, and implementations has a large practical impact and can reduce the carbon footprint of the computation. We are particularly interested in the development of improved simulation methods for electronic structure computation that targets heterogeneous computer architectures comprising GPUs and FPGAs. By exploiting massive parallelism, mixed and low-precision computing, and hardware customization and operating each architecture at its sweet spot, substantial gains in efficiency and performance can be achieved. Recent example of this co-design approach is our highly scalable and efficient submatrix method (<https://doi.org/10.1016/j.parco.2022.102920>) or our FPGA-accelerated computation of electronic repulsion integrals for quantum chemistry (<https://doi.org/10.48550/arXiv.2303.13632>), which are being integrated in CP2K.

Examples for PhD projects in this area are: The development of novel, highly scalable and efficient methods for atomistic simulations with DFT-variants on modern hardware accelerators, in particular with a combination of GPUs and FPGAs; investigation of approximate computing methods that combine low and high-precision arithmetic to reduce time/energy to solution while limiting the effect on the quality of results.

### **4#PC2: Methods for compressing smooth volumetric data on grids**

Since a few decades, modern experimental and computational science is more and more confronted with the "big data" issue, i.e. the necessity to handle - and possibly store - huge amounts of intermediate raw data in order to obtain the desired results. One strategy to mitigate such issues is to compress the raw data in a suitable way in order to reduce bandwidth and storage requirements. The efficiency of a compression algorithm depends a lot on the precise nature of the data. If the general structure of the input data is well known, tailor-suited compression algorithms can be designed which exploit that particular structure. We are developing such tailor-suited approaches to compress time series of volumetric data on grids which are smooth both in space and time. For the application of storing total electron density data from ab initio calculations, a lossless compression ratio of 40 : 1 can be reached (<https://pubs.acs.org/doi/10.1021/acs.jcim.8b00501>). However, for other categories of volumetric data such as wave functions or molecular orbitals, a slightly different fine tuning will be required to achieve a similar compression efficiency.

Examples for PhD projects in this area: Designing compression algorithms for other types of volumetric data time series; Optimizing the compression and decompression code for more efficiency; Parallelization of the compression and decompression operations.

## **5#PC2: Electronic Structure Theory on Quantum Computers**

The reduced density matrix functional theory method (RDMFT) is a post-DFT method that can reach a computational accuracy well beyond conventional DFT. While the required reduced density matrix functional can be approximated similarly to approximate DFT functionals like LDA or GGA, the big advantage of RDMFT over DFT lies in the fact that the use of the one-particle reduced density matrix opens the door for novel approximations in which the computational bottleneck is the evaluation of quantum-mechanical expectation values. To address this challenge, we have worked on a hybrid implementation where the wave function is optimized on a conventional digital computer but the most demanding part of evaluating quantum-mechanical expectation values is evaluated on a quantum computer (<https://doi.org/10.1103/PhysRevResearch.4.033160>).

Possible topics for a PhD project in this area may involve: Devising and implementing hardware-efficient trial-states; development of methods to reduce the number of quantum programs; integration in Qiskit environment; practical evaluation on quantum computers; interfaces between common quantum chemistry codes and quantum computers; development of other novel approaches for quantum chemistry on quantum computers.

TU Dresden is one of the largest technical universities in Germany. It has a long history of running HPC installations and has been specialized for data intensive HPC for a long time. The focus topics of NHR@TUD are (1) Methods for Big Data and data analysis as well as data management, (2) Machine Learning, (3) Tiered storage architectures and I/O optimization, and (4) Performance and energy efficiency analysis and optimization. These method developments focus mainly on the application areas life sciences and earth system science. The NHR center at TU Dresden cooperates closely with the the AI center ScaDS.AI Dresden/Leipzig (Center for Scalable Data Analytics and Artificial Intelligence).

## **1#TUD: Software Tools for Parallel Performance Analysis and Optimization**

Parallel performance, efficiency, and scalability are primary concerns for the HPC community. Dedicated software tools for performance analysis and optimization are indispensable to cope with a number of challenges for HPC programmers and cluster operators.

There are several performance aspects waiting for new software tool approaches for measuring, analyzing, and steering parallel performance, among others: new heterogeneous hardware architectures; new and evolving parallel programming models; more and more complex application codes; Big Data and Machine learning frameworks in HPC environments; AI analysis of monitoring data; as well as high scalability and low overhead of the tools themselves.

This topic offers the opportunity to join an established performance tools research group working on well-known software tools renowned in the HPC community.

## **2#TUD: Task Parallel Programming Models**

Task based parallel programming models are a promising alternative to message passing and multi-threading. There are a number of task parallelization models with a strong HPC focus such as OpenMP, Cilk, Intel TBB, StarPU, Chapel, X10, HPX, Charm++, and a number of research approaches. They are rather diverse in their shared-memory/distributed memory scopes, parallel scale, task granularity, and advanced features like built-in faulttolerance.

Current challenges for their adoption in Computational Science projects are (gradual) porting of existing applications to task parallel models and interfacing between task based and conventional software components. Challenges from the Computer Science perspective are task scheduling and migration strategies, support for heterogeneous hardware architectures, fault tolerance approaches, and the integration into programming language standards.

## **3#TUD: Data-Driven HPC Workflows**

Work on new concepts for data-driven workflows that connect HPC with cloud computing and edge computing where each processing step is executed in the best-suited environment. This may combine methods from the different scopes which offer mutual advantages. For example, combine HPC technologies like Infiniband, RDMA, NVMeoF, GPU direct and others with cloud protocols like S3, RabbitMQ, ZeroMQ, MQTT, Apache Kafka, and others. This may also involve workflow management approaches compatible with the respective HPC, cloud, or edge counterparts. It may also focus on data and metadata management integration in data-driven workflows. Finally, it may concern performance and scalability improvements for the most demanding processing stages.

#### **4#TUD: Machine Learning**

After their introduction and strong further development, machine learning methods are used these days in many scientific application areas. In particular, modern HPC infrastructures provide the necessary large-scale computing power to solve complex learning approaches based on large amounts of data. In this context, the efficient use of resources is a current field of research, in order to make the training of large learning models not only possible, but also efficient at the same time. In addition to the investigation and optimization of network architectures, this also involves the efficient distribution strategy of the data as the basis of an efficient learning process. Furthermore, current research is conducted to establish services for the optimization of the usually large hyperparameter space of aforementioned applications. Within the national AI competence center ScaDS.AI Dresden/Leipzig, ZIH and its partners are involved in research on efficient distributed learning approaches and performance optimization of machine learning applications on scalable HPC infrastructures and also on alternative computing architectures for the future.

#### **5#TUD: Energy Efficient Computing**

Limited power and energy budgets have become major challenges in many areas of (scientific) computing. ZIH has developed a strong expertise in power and energy measurements, addressing important aspects such as accuracy, scalability as well as temporal and spatial granularity. This work is complemented by the development of methods such as modeling and optimizing the energy efficiency of parallel applications. Benefiting from several decades of work in performance benchmarking processor and memory architectures, we are developing a growing body of Open Source software tools to address optimization challenges, e.g., BenchIT for low-level benchmarking, lo2s for extremely lightweight node-level energy and performance monitoring, and MetricQ for highly-scalable, distributed metric data processing. We also consider system cooling to be an integral part of the energy efficiency optimization, and waste heat reuse as a means to mitigate unavoidable detrimental effects on the environment. As such, our approach to energy efficiency research is truly holistic, spanning from the single processor core to the data center. And you can contribute!

#### **6#TUD: Machine Learning for Research Data Management**

Data management (DM) is often perceived as an additional effort detaining researchers from their desired job, doing research. Nevertheless, good scientific practice and traceability of research outcomes demand DM throughout the research process. Especially, working with large amounts of data requires enhanced techniques to keep DM tasks in check, streamlined to the research process. Automatic detection of discrete datasets within a data collection and its significant metadata may reduce individual DM efforts significantly. However, depending on the precise research question, similar data could be handled and analyzed very differently. Machine learning algorithms could help the researcher to structure large amounts of data and perform important DM tasks in time, after a learning phase, e.g., on discrete dataset definition, important metadata, and project schedule. Ideally, researchers can hand over large parts of time-consuming DM tasks after a learning phase and become DM supervisors rather than performing DM tasks themselves. This will allow attractive PhD developments in an important and strategically growing scientific field.

## NHR@ZIB – Berlin

At Zuse Institute Berlin (ZIB), model-based simulation and optimization, data-driven problem solving and advanced computing meet with high-performance research services and transfer to diverse application fields and to industry. The NHR center at ZIB is part of a research institute with a focus on Scientific Computing and a broad spectrum of interdisciplinary research projects and about 300 scientists and research infrastructure specialists. The NHR center at ZIB will take the leadership in selected application domains, e.g., life sciences with focus on model-driven simulations, advanced integration of artificial intelligence / machine learning and simulation, or heterogeneous HPC architectures jointly with other NHR centers.

### **1#ZIB: Hybrid Quantum-Classical Algorithms for Combinatorial Optimization**

Quantum computing has the potential to revolutionize HPC in the long term. At ZIB, we aim at developing a workflow from the design of hybrid quantum-classical algorithms up to their simulation by means of quantum simulators on HPC hardware. We consider this specifically in the context of combinatorial optimization problems, because of the industrial relevance as well as first promising results from quantum computing for this problem class. The problem comprises three sub-aspects: (1) Identification of combinatorial optimization problems that can be solved using NISQ and QAOA in combination with traditional approaches, (2) development and design of hybrid quantum-classical approaches for selected examples of the problems identified in (1), and (3) simulation and verification of quantum-classical hybrid algorithms. Since we do not expect that general-purpose quantum computing will be available in the near future and with sufficient scaling, a preliminary verification by quantum computing simulators running on classical HPC hardware is essential.

### **2#ZIB: Large-scale Explainable AI**

AI has the potential to revolutionize how we do science and it will be at the forefront of the next wave of innovation. In particular methods capable of learning from large-scale combined with methods from explainable AI (XAI) will play an important role in future applications. This requires new methodological approaches, such as e.g., new optimization methods and more generally new learning algorithms. At the same time, it requires new approaches to handling and processing large-scale data. At ZIB we develop both the underlying methodologies for large-scale learning and XAI as well as well the necessary HPC approaches to effectively execute these approaches. This includes the full design stack from high-level end-to-end learning methodologies to lowest level efficient data handling procedures.

### **3#ZIB: Semantic I/O with Object Stores in HPC**

Object stores like S3 are a common storage infrastructure in cloud environments. Nowadays, object stores become increasingly popular in HPC as an alternative concept to store data as they overcome scalability issues of POSIX-complaint file systems used traditionally in HPC configurations. Examples of HPC-relevant objects store are based on distributed key/value stores. This offers the possibility to integrate application-specific metadata directly into the managed objects at the low level. Moreover, with the flexibility of a key/value store the efficiency of I/O operations can be further increased as the logical structure of the data to be stored can be mapped to an optimized schema of the key/value backend. As such, these data containers are natively supported by the storage system with best performance for storing and retrieving the data. At ZIB, we perform research with one of the current implementations of a distributed object store, the Distributed Asynchronous Object Store (DAOS). Using its infrastructure, the potential benefit of taking the semantic knowledge of data into account will be investigated.



#### **4#ZIB: Modelling Excitations in Novel Materials**

Ab-initio calculations are a valuable tool to study materials for new optoelectronic devices, including those that exploit solar energy. For an accurate description of optical and electronic properties, standard density functional theory (DFT) approaches are not sufficient. Methods going beyond DFT, especially those based on the GW approximation and the Bethe-Salpeter equation, or time-dependent DFT methods, are required instead. Their application, however, involves larger computational resources. The PhD candidate will take advantage of powerful hardware available at NHR@ZIB to implement and apply these methods on modern energy efficient architectures. Another closely related research topic is the use of the aforementioned methods to study the dynamics of exciton creation and propagation in materials used in optoelectronic devices. Depending on the research direction taken, these topics may be combined with the development of tools that expand the functionality of a selected ab-initio package.

#### **5#ZIB: Cross-Platform and AI-enhanced Molecular Dynamics**

The convergence and combination of classic HPC methods with approaches from artificial intelligence (AI) is an ongoing trend in science. Molecular dynamics (MD) benefits from AI as well: High-dimensional neural networks potentials enable the accurate modeling of many-body interaction in materials research. They allow for predictions of undiscovered chemical species, describing the chemical space of a given structure, or precisely predict the pathways between folded and unfolded conformations of a protein. To effectively support molecular dynamics research with these promising AI-based models, they need to be combined with existing HPC methods and codes. Also, accelerator hardware is required to fulfill the computational demands of such research. ZIB is working closely with researchers and software developers from the MD domain to enable such combinations of AI and HPC. In addition, a focus of ZIB's research in the supercomputing department is on the cross-platform support of AI methods on accelerators by using modern programming languages and frameworks.

#### **6#ZIB: Portable and Energy-Efficient Computing on Data-Parallel Platforms**

Today, with GPUs, TPUs, FPGAs and others, a variety of data-parallel accelerators with native support of multiple data representations are available. The more domain-specific algorithms can be effectively mapped on these accelerators, the more an energy-efficient data processing can be realized in the future. Whereas for dense and partially sparse linear algebra advantages of mixed precision and approximate computing strategies has been demonstrated on these devices, there are a large number of questions in other areas of numerical mathematics like PDE solvers unanswered. Further, from the technical point of view some challenges need to be tackled. The current "art of coding" is under question as with the diversity of accelerators one has to select the associated programming models. At ZIB, we want to continue our interdisciplinary work in areas of mathematical numerics and modern software engineering by exploring the potential of mixed-precision algorithms, e.g., on unstructured grids on current and next-generation data-parallel processors. For that, our emphasis are programming models portable across vendors and processor architectures (e.g., oneAPI).

## **7#ZIB: Massive Simulation of Agent-Based Models**

Agent-based models (ABMs) have become very popular for describing the interaction of a large number of similar entities/agents (human beings, households, institutions or the like) by a set of rules describing the changing location of agents and their interaction with each other. Interacting agents can alter their own or the other agent's properties, for example, by changing an opinion or transmitting an infection. Information about agents' mobility and status dynamics is typically deduced from available data sets and expert knowledge. Due to their versatility, ABMs are very attractive for modeling spatiotemporal population dynamics in various disciplines, ranging from social process models via traffic or mobility transition models to opinion spreading and infection dynamics. ABMs with really large numbers of agents (millions or larger) require high-performance computing infrastructure for the required massive simulations, especially if scenario studies with changing parameters or different external control are investigated. AT ZIB, we are developing tools for such massive ABM simulations on HPC systems.