

NHR- Graduate School: Research Fields

List of Contents

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

ZIH – NHR@TUD – Dresden	13
1#TUD: <i>Software Tools for Parallel Performance Analysis and Optimization</i>	13
2#TUD: <i>Task Parallel Programming Models</i>	13
3#TUD: <i>Data-Driven HPC Workflows</i>	13
4#TUD: <i>Machine Learning</i>	14
5#TUD: <i>Energy Efficient Computing</i>	14
6#TUD: <i>Machine Learning for Research Data Management</i>	14
NHR@ZIB – Berlin	15
1#ZIB: <i>High-Performance Agent-Based Simulation Environment</i>	15
2#ZIB: <i>Hybrid Quantum-Classical Algorithms for Combinatorial Optimization</i>	15
3#ZIB: <i>Large-scale Explainable AI</i>	15
4#ZIB: <i>Semantic I/O with Object Stores in HPC</i>	16
5#ZIB: <i>Portable and Energy-Efficient Computing on Data-Parallel Platforms</i>	16
NHR SW	17
1#NHRSW: <i>Algorithmic Differentiation in HPC environments (TU KL/ Uni Saarland)</i>	17
2#NHRSW: <i>Compilation Techniques and Domain-Specific Languages for Heterogeneous Systems (Uni Saarland)</i>	17
3#NHRSW: <i>Data Augmentation with Rule Based Systems (Uni Saarland)</i>	17
4#NHRSW: <i>Multiscale modeling in theoretical condensed matter physics (JGU Mainz)</i>	18
5#NHRSW: <i>Performance optimization of life science applications (JGU Mainz)</i>	18
6#NHRSW <i>Data-intensive HPC and machine learning applications in hadron physics (JGU Mainz)</i>	18
7#NHRSW: <i>Lattice QCD codes on accelerators and Machine Learning techniques in lattice field theory (JGU Mainz)</i>	19

NHR4CES@RWTH – Aachen / 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 Rude)

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: Efficient numerical and parallel methods for scale bridging

Multi-scale numerical simulations require to capture effects on different spatial, temporal, or other scales. A prime example is the modeling of droplets which requires to capture the nanoscale physics as well as the mesoscale resolution of the droplet shape and interaction of multiple droplets. The uniform numerical treatment and resolution of a multi-scale problem is a great computational challenge and very well-suited for HPC systems. It requires the combination of advanced numerical approaches and scalable parallel algorithms to solve the typically coupled nonlinear partial differential equations. In every multi-scale application, the ultimate goal is to establish a systematically hierarchical approach and uniform simulation framework, including programming interfaces for data progressing and including numerical approaches in combination with a consistently parallel solution with high scalability properties.

2#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.

3#KIT: Data-intensive computing

Enormous amounts of data are produced by ever higher scaling simulations on HPC systems, but also by large-scale experiments. In order to analyze this data and extract new scientific findings from it, services that accompany the appropriate infrastructure are required to process and evaluate such data. Creating an excellent environment for the management and analysis of research data combined with informatics and mathematics R&D is an important topic, given the rapid growth rates of real and simulated data, the growing complexity of information systems, and the prospects of progress in the entire field of data analytics using AI. The objective of NHR@KIT is to achieve a deep understanding of the information content of experimental and simulation data, and to create data-driven models, thus transforming data into knowledge.

4#KIT: High performance computations in materials science and climate research

New and improved products demand highly tailored materials properties, which are influenced not only by the chemical composition of the material but also its microstructure. Hence determining the microstructural evolution, especially for new materials and processing ways, is of great importance. The phase-field method will be used to investigate the microstructural evolution during the sintering process, which is both industrially important and of interest to climate models since glaciers are in a continuous state of sintering. As sintering requires resolving particles on the nanometer to micrometer scale as well as a sufficiently large number of particles to be statistically representative, an efficient and scalable solver needs to be developed and applied to the problem. Since sintering requires long-range motion, special communication patterns need to be implemented and optimized. Due to large simulation volumes, the resulting data needs to be analyzed in an efficient, parallel framework which exploits modern data analysis libraries for rapid development of toolchains.

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.

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. (Prof. Christian Plessl)

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. (Prof. Christian Plessl)

3#PC2: Methods for Atomistic Simulations on Modern 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 wide applicability and numerical demand of these methods, they contribute a very substantial portion of the global workload of HPC systems. Any improvement in efficiency in the underlying methods, algorithms, and implementations will thus make an important contribution to reduce time to solution and the carbon footprint of the computation. We are particularly interested in the co-design of improved simulation methods for linear scaling density functional theory (DFT) that are custom-tailored to effectively exploit the properties of modern computer architectures, that is, massive parallelism, vectorized execution, mixed and low-precision computing. If these capabilities are properly exploited, performance or efficiency improvements of orders of magnitude are possible. A recent example of this co-design approach is our highly scalable and efficient submatrix method (<https://doi.org/10.1016/j.parco.2022.102920>) that is 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 GPUs or 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. (Prof. Thomas D. Kühne)

4#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://arxiv.org/abs/2202.02417>).

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. (Prof. Thomas D. Kühne)

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. Its NHR center's focus topics are (1) Methods for Big Data and data analysis as well as management, (2) Machine Learning, (3) Tiered storage architectures and I/O optimization, and (4) Performance and energy efficiency analysis and optimization. The application area focus topics are (5) life sciences and (6) earth system science. The NHR center at TU Dresden cooperates closely with the 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 alternatives 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 fault tolerance.

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.

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.

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: High-Performance Agent-Based Simulation Environment

Complex systems composed of many interacting units are ubiquitous in today's world. Complex systems sciences aim to understand the behavior, movement, and dynamics of individual units, and also how macroscopic system properties emerge from their interaction. The units may be individuals or organizations in the social sciences and humanities, microscopic particles, such as molecules, colloids or cells in the natural sciences, market agents, such as retail customers in finance and economics, or sensors, autonomous vehicles in engineering. A common approach to analyzing the dynamics of such complex systems are many-particle or many-agent models. The computational cost of agent-based models (ABMs) is very high and for stochastic ABMs a lot of individual simulations are required to sample quantities of interest. In many applications, e.g., in simulations of pandemic infection spreading, ABMs use very large numbers of agents (many millions). At ZIB, research on ABMs is oriented towards building a framework designed to create, execute and analyze agent-based simulations in HPC environments with the goal to fully utilize the resources available on current heterogeneous platforms consisting of multi-core CPUs and many-core GPUs. The interested PhD candidate would become member of a larger interdisciplinary team and on high-performance agent-based simulation with a broad spectrum of specific research topics to choose from.

2#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.

3#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.

4#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.

5#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).

NHR SW

1#NHRSW: Algorithmic Differentiation in HPC environments

(TU KL/ Uni Saarland)

The landscape of HPC architectures, special accelerators and communications models has evolved over the last decades and will continue to do so in the future. This implies, that new libraries and programming models are used to fully facilitate the peak performance of the new systems. For all these architectures and libraries, it is an ongoing challenge to provide automatically created derivative code via the theory of Algorithmic Differentiation (AD). Examples are the PGAS programming model, CUDA, ROC, FPGAs, TPUs or GPUs. Each library needs to be analyzed for its specific challenges and extended such that a differentiated version can be used in the context of AD.

Another aspect is the formulation of problems in domain-specific languages (DSLs) which makes the formulation more concise. Furthermore, the generated code can be optimized for different hardware architectures and run-time environments. Formulating problems also requires the use of derivatives or the sensitivities for the full problem formulation are required in a design optimization context. Algorithmic Differentiation (AD) techniques can be used to enhance DSL frameworks with the capability to additionally generate the derivative code of the problem formulation. The generated derivative code is then again optimized for different hardware architectures and run-time environments. (Prof. Nicolas Gauger, Prof. Sebastian Hack)

2#NHRSW: Compilation Techniques and Domain-Specific Languages for Heterogeneous Systems (Uni Saarland)

Future HPC systems will become more and more heterogeneous including accelerators of various kinds. GPUs are already widely adopted but in the future more specialized accelerators such as TPUs or FPGAs could make their way into HPC systems because they promise better energy efficiency. However, hardware heterogeneity poses a significant challenge to programming these systems because it deviates from standard programming models. Our group does basic research on domain-specific languages (DSLs) and compiler techniques to enable the programming of future heterogeneous systems. We work on supporting next-generation vector instruction sets within the LLVM compiler toolchain and on the DSL framework AnyDSL that allows for simple and effective DSL embedding using partial evaluation and compilation to a wide range of accelerators including NVIDIA and AMD GPUs, FPGAs, CPUs with various vector extensions. (Prof. Sebastian Hack)

3#NHRSW: Data Augmentation with Rule Based Systems (Uni Saarland)

In many applications, there is a bottleneck in training data. Only in exceptional cases like language models for English, enough data exists to train the huge model. In many other situations not enough data is available to train large scale neural networks. However often rules based systems (e.g. grammars) exist that can not only be used to analyse data but also to generate new data. The goal is to explore neural network training algorithms that can utilize data generated from rules based systems. (Prof. Dietrich Klakow)

4#NHRSW: Multiscale modeling in theoretical condensed matter physics (JGU Mainz)

Multiscale modeling is a central topic in theoretical condensed matter physics and materials science. One prominent class of materials, whose properties can rarely be understood on one length scale and one time scale alone, is soft matter. Therefore, several ideas have been proposed how to carry out simulations of soft materials at several levels of resolution in one simulation. One important issue is to ensure that such multiresolution studies are more efficient by orders of magnitude than highly optimized molecular dynamics simulations at the finest required level of resolution. This is challenging, because different types of data structures must be combined in an efficient manner. Applying established optimization and parallelization methods such as geometric domain decomposition becomes highly challenging in a multiresolution context. The goal of the thesis will be to develop new solutions for such problems. Possible approaches are to develop adaptive load-balancing schemes for molecular dynamics simulations, allow applications to steer their IO-behavior and provide an asynchronous checkpointing environment, which is able to scale by distributing the checkpoints in the HPC cluster. (Prof. Brinkmann)

5#NHRSW: Performance optimization of life science applications (JGU Mainz)

Recent years have seen a tremendous increase in the volume of data generated in the life sciences, especially propelled by the rapid progress of next generation sequencing (NGS) technologies. High-throughput sequencers can produce massive amounts of short DNA strings (called reads) in a single run. This leads to large-scale datasets being processed in a wide range of bioinformatics applications. Furthermore, the cost of these technologies has been decreasing dramatically. The low sequencing cost per genome enables widespread usage and renders population-scale projects feasible. Examples include the Earth BioGenome Project, metagenomics microbiome sequencing studies, and world-wide SARS-CoV-2 sequencing efforts. However, the analysis of large sequencing datasets poses hard computational challenges. In particular, read mapping, read classification, and de-novo genome assembly are performance-critical tasks, being initial stages required for manifold types of NGS analysis pipelines. Parallelization can be employed to reduce execution times but imposes additional challenges due to large storage requirements of index data structures, and their associated irregular memory access patterns. In summary, algorithmic design and implementations of bioinformatics applications struggle to keep pace with recent high-throughput sequencing techniques and their ever increasing data acquisition rates. This theme thus focuses on the design, implementation and evaluation of scalable algorithms, data structures, and tools for life science. Our approach will be based on using modern high performance computing (HPC) technologies (such as GPUs and GPU clusters) in order to design massively parallel algorithms and data structures that serve as a foundation for a wide variety of tools such as de-novo genome assemblers, read aligners, metagenomics, and transcriptomics. Our methods and tools are often developed in collaboration with interdisciplinary partners between Computer Science and the Department of Biology, the Mainz Medical School, or TRON. (Prof. Bertil Schmidt)

6#NHRSW Data-intensive HPC and machine learning applications in hadron physics (JGU Mainz)

Many observables in hadron physics require the identification of very rare signals. Hadron physics facilities like for example the future international Facility for Antiproton and Ion

Research under construction at GSI in Darmstadt will allow for the investigation of electromagnetic structure observables in an environment of a 10^6 to 10^7 times higher hadronic background. Examples are electromagnetic form factors of the nucleon and Transition Distribution Amplitudes. The identification of these rare signals require an efficient usage of the underlying storage media in order to improve the data throughput. Moreover efficient analysis of the data require further development of machine learning algorithms and application of AI to identify the signal from the background. The algorithms will be applied to pseudo-data of the future antiproton annihilation experiment PANDA at FAIR as well as to real data of the BES-III experiment running at the BEPC-II collider in Beijing. (Prof. Frank Maas)

7#NHRSW: Lattice QCD codes on accelerators and Machine Learning techniques in lattice field theory (JGU Mainz)

The formulation of Quantum Chromodynamics (QCD) on a four-dimensional space-time lattice allows for an ab initio treatment of the strong interaction. Markov Chain Monte Carlo techniques are used to compute physical observables numerically. Typical calculations are performed on systems involving 10^8 grid points, which requires a massively parallel approach to computing, as well as state-of-the-art algorithmic methods, including efficient solvers, various kinds of preconditioning and deflation techniques and efficient implementations for the calculation of correlation functions ("contractions"). In order to benefit from new computing hardware, these codes must be ported and optimised to architectures such as GPUs to guarantee high performance and good scaling properties.

A related field is, that the exploring of the potential of machine learning techniques for applications in lattice field theory has turned into an active field of research. Our project is focussed on the question whether ML can serve as a viable alternative to standard Markov Chain Monte Carlo. One line of research pursues the efficiency of ML for two-dimensional spin models compared to standard algorithms such as cluster algorithms or multigrid techniques. Another important application concerns the applicability of ML to the problem of determining observables. Furthermore, we will perform a feasibility study into training networks to infer physical observables for different system size. In this way one may cover a wider range of system parameters, which could result in huge computational savings. (Prof. Wittig)