NHR- Graduate School: Research Fields

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 (RWTH/TUDa)

connects NHR4CES with the National Research Data Infrastructure (NFDI) initiatives and supports users at the interface to HPC and emerging solutions for distributed research data management.

2#NHR4CES: Data Science and Machine Learning (RWTH/TUDa)

provides methodological support across application domains and guidance to make HPC and data infrastructure better suited for new tasks

3#NHR4CES: Parallelism and Performance (RWTH/TUDa)

support for performance engineering, in particular performance management, modeling, optimization, and parallel programming

4#NHR4CES: Visualization (RWTH)

solutions for interactive (optionally immersive) visualization of scientific datasets

5#NHR4CES: Energy Conversion (RWTH/TUDa)

enables computationally efficient simulations of real-scale combustion devices by developing HPC-ready reactive CFD (rCFD) software and methods by a co-design process

6#NHR4CES: Fluids (RWTH/TUDa)

advances the use of HPC to enable highly resolved and sophisticated CFD applications

7#NHR4CES: Materials Design (RWTH/TUDa)

provides the best tools for the analysis and visualization of computed 3D particle configurations

8#NHR4CES: Digital Patient (RWTH)

model-based medical applications, ranging from molecular modeling of disease mechanisms up to patient scale system medicine models

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, and ligand-induced signalling. (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)

NHR@GÖTTINGEN – GWDG, Universität 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, synergetic 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 data-driven 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: Digital Twins for Data-Centers

Data centers contain various infrastructure to, e.g., steer airflow and cooling of housed systems. As part of this topic, we aim to create a 3D model of the data center including the technical processes and relevant characteristics. This allows investigating the impact of various tuning options and together with 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.

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://arxiv.org/abs/2104.08245) 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-mechanicalexpectation values is evaluated on a quantum computer.

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)

5#PC2: Machine Learning Based Optimization of Photonic Links

Optical computers based on integrated photonic waveguides promise strongly desirable advantages such as high speed and low energy consumption/heating. Also, they might provide a platform for scalable optics-based quantum computing. However, the search for a single material providing all the required constituents including light sources, low-loss passive lines, and nonlinear transistor-like elements in sufficient quality has not been successful so far. One approach to deal with this is to connect or sandwich chiplets of different materials. This requires highly efficient inter-chip communication, either directly via free space or via optical fibers.

In our previous research on optical antennas driven by colloidal quantum dot light sources we already achieved directive emission much higher than previous designs using our simulation-driven approach. Based on this experience with optical antennas and integrated photonic waveguides we want to develop novel concepts for optical chip-to-chip (fine-tuned single antennas, phased antennas, MIMO-type) and chip-to-fiber transmission by HPC simulation-based optimization, using classical nonlinear optimization and machine-learning supported approaches. (Prof. Jens Förstner)

NHR@TUD-ZIH

TU Dresden is one of the largest technical universities in Germany. The Center for Information Services and HPC (ZIH) at TU Dresden is the university IT center and focus in the NHR context on (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, and the application areas (5) life sciences and (6) earth system science. The NHR center at TU Dresden cooperates very closely with the ScaDS.AI (Center for Scalable Data Analytics and Artificial Intelligence) Dresden/Leipzig.

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

Parallel performance, efficiency, and scalability are prime criterions for the HPC community. This is true from both perspectives, the computational science perspective which requires faster results as well as the capability to tackle very large problems at all, as well as from the computing center perspective which pays attention to cost efficiency and energy efficiency. Dedicated software tools for performance analysis and optimization are indispensable to cope with a number of challenges for HPC programmers.

First, more and more complex heterogeneous hardware architectures. Second, new and evolving parallel programming models. Third, more and more complex application codes combining various hardware and software aspects. Last but not least, performance tools need to be scalable and efficient themselves and pay very close attention to the run-time overhead they cause onto the target applications. Current challenges are (1) fine-grained and highly scalable monitoring of task parallel programming models and (2) monitoring for various Big Data and Machine Learning frameworks and their interpretation towards optimization potentials.

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: New Approaches for Data-Intensive and Data-Driven HPC

For almost two decades, data-intensive and data-driven HPC has been a focus topic at ZIH. Several generations of HPC installations have been designed with particular attention to data storages and data transfers.

Besides the conventional usage patterns for HPC storages (parallel file systems) and data transfers (IB and RDMA), there are promising alternatives from the HPC area (NVMEoF) and from outside the HPC community. The latter include object storages (S3 protocol, MinIO server) and messaging/streaming systems and message brokers (RabbitMQ, ZeroMQ, MQTT, Apache Kafka, and others). They offer larger flexibility and robustness compared to native HPC approaches and promise much better integration with Cloud-style data-driven workflows, where HPC is an essential step but not the only processing step. Open challenges are better integration of such technologies with HPC infrastructure, adaption to HPC software environments, interoperability with workload managers (Slurm), compatibility with or even replacement for parallel programming models, porting of existing HPC applications toward such technologies, and performance/scalability optimization as part of HPC applications.

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. From system design, to operation, to optimization – performance and energy efficiency are intrinsically linked and, therefore, need to be conjointly considered. For more than ten years, ZIH has developed a strong expertise in power and energy measurements, addressing important challenges such as accuracy, scalability as well as temporal and spatial granularity. This work is complemented by the development of methods such as modelling and optimizing the energy efficiency of parallel applications. Benefitting from several decades of work in 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 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.

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 if 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).