COMPETENCE CENTER HIGH PERFORMANCE COMPUTING

The Chair of Scientific Computing at the University of Kaiserslautern – partner at the Performance Center Simulation- and Software-Based Innovation – and CC HPC are working on behalf of MTU Aero Engines AG to improve and accelerate the design software for engine optimization. The core of our contribution is the software GPI-Space, that helps to realize "Memory Driven Computing" architectures.

DR. FRANZ-JOSEF PFREUNDT
HEAD OF DEPARTMENT

INNOVATION, DISRUPTION AND INTEGRATIVE THINKING IN THE WORLD OF DISTRIBUTED COMPUTING

The department has developed innovative world-class technologies for solving large data problems, specifically BeeGFS, Pre-Stack PRO and the Global Address Space Programming model (GPI) in addition to the Big Data framework called GPI-Space. In recent years, we have gained international attention by successfully combining these technologies with deep learning methods. At its core, it's always about the scalable automatic parallelization of big data problems. This is based on the concept of "Memory Driven Computing" which combines scalability and performance. In the ALOMA and SafeClouds projects, we are further developing industry-specific solutions based on this technology.

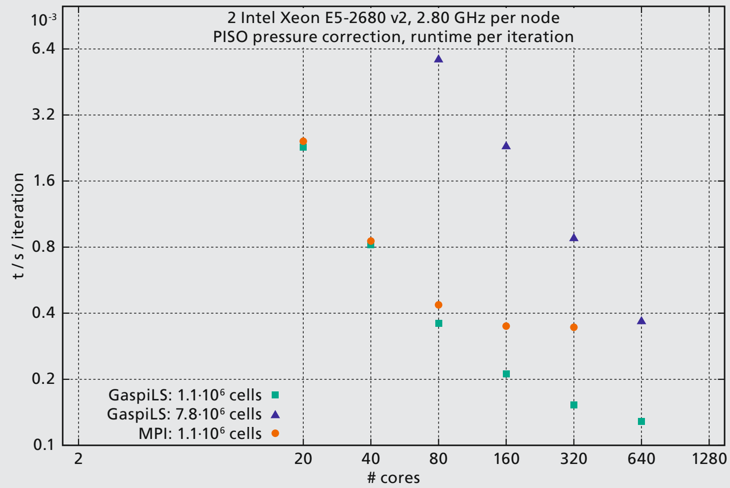
The aim of our involvement in EU-sponsored HPC research is to strengthen European technologies and improve the marketability of European HPC software products. In addition, our goal is to bring together microelectronics development and application development in co-design projects. In the application-specific development of computer hardware we see a way to improve Europe's position in the fast-growing market for HPC / Big Data solutions.

The energy systems of the future will consist of millions of distributed IoT computers. These are used to optimize the self-consumption of PV power, regulate the creation of community grids, control large and small power storage systems and coordinate the energy flow in our power grids. Within our projects technologies and solutions are developed to master this distributed computing world. We are committed to find intelligent solutions that work to advance the energy transition.

Contact
franz-josef.pfreundt@itwm.fraunhofer.de
www.itwm.fraunhofer.de/en/hpc

MAIN TOPICS

- Green by IT
- BeeGFS – Parallel Cluster File System
- Visualization
- Seismic Imaging
- Data Analysis and Machine Learning
- Scalable Parallel Programming



1

GaspiLS – SCALABILITY FOR CFD AND FEM SIMULATIONS

1 Pressure correction computation: GaspiLS (green and blue) has significantly improved performance and scalability in comparison to the MPI based implementation (orange).

Many simulations in engineering are based on CFD and FEM methods. Examples are the determination of aerodynamic properties of planes or the analysis of the statics of buildings. The vast majority of computation time is spent in the solution of the underlying equations. The performance of the employed iterative solvers has a significant impact on the total run time of these simulations. As such, they also significantly impact the insight which can be generated by those simulations.

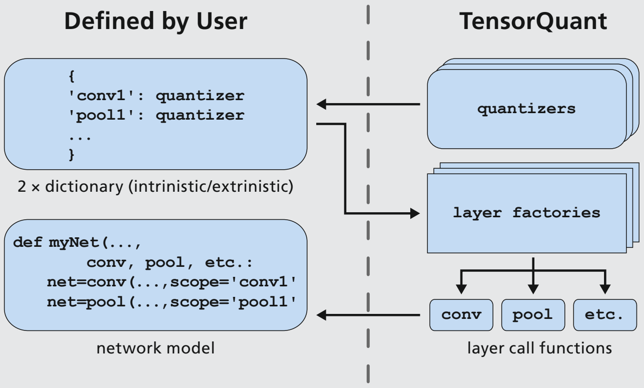
We have developed the linear solver library GaspiLS to gain more insight from the simulations.

Industry uses GaspiLS for better scalability

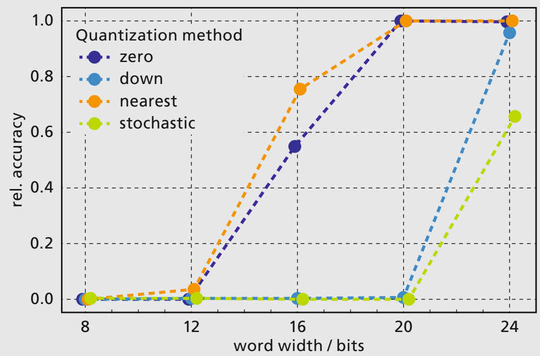
Scalability measures the parallel efficiency of an implementation. The optimum is the so-called linear scalability. This corresponds to a full utilization of the cores within a single CPU or the CPUs within a cluster, which are interconnected by a network. A better scalability allows to use the compute resources more efficiently which implies a shorter time to gain a solution. Ultimately, better scalability allows for more detailed models, more precise parameter studies and a more cost efficient resource utilization.

Better compute resource utilization

In order to achieve better scalability, GaspiLS uses tools for parallel programming which are developed by our group; these are the communication library GPI-2 and its underlying programming model. The algorithm is split into fine grained sub problems (so-called tasks) with mutual dependencies. This allows for the assignment of executable tasks to free compute resources at any time and guarantees for a continuous stream of compute tasks for every CPU. The avoidance of global synchronization points and the huge amount of generated sub problems allows to hide potential communication latencies and to compensate for imbalances in the compute time. Every single core is maximally emploited at any time.



1



2

TENSORQUANT BRINGS DEEP LEARNING TO MOBILE APPLICATIONS

Machine learning methods are being used more and more in the industrial and service sector. Especially artificial neural networks or Deep Learning (DL) have a high impact on the development of intelligent systems. Research continuously provides new Deep Learning methods, which open a wide range of possibilities for these algorithms in many different practical application scenarios.

However, a significant technological hurdle on the way to such applications in production is the enormous computational effort required to calculate and evaluate the DL models.

This explains why the development of specialized DL hardware has recently come into focus. In the future, new chip and memory architectures will enable the use of high performance hardware components that save energy and, at the same time, expand the use of DL, for example to autonomous vehicles, mobile phones, or integrated production controls.

Learning does not require high precision in numerical processing

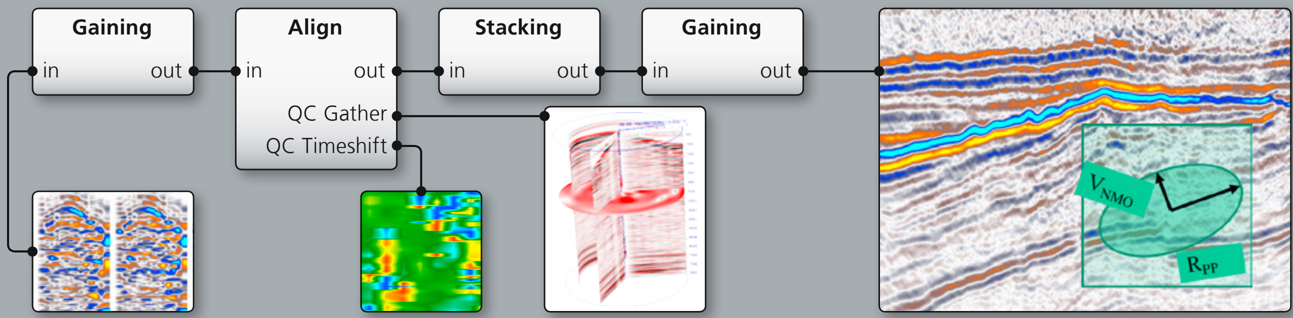
We exploit a mostly mathematical feature of DL: Learning and evaluating models can be reduced to a numerical computation of a small number of operations using tensor-algebra (for example, matrix-multiplications). In addition, tensor calculation works well with much less precision in terms of numerical processing than it is typically the case with physical simulations. In comparison to general computational units such as CPUs and GPUs, these features enable a highly efficient hardware implementation.

TensorQuant allows the emulation of machine learning hardware

In the development of DL applications on specialized hardware, difficulties are encountered as the minimum requirements for computational precision vary significantly between the individual models. As a result, the simultaneous optimization of DL models and hardware is difficult in terms of computational performance, power consumption, and predictive accuracy. Our TensorQuant (TQ) software lets developers identify critical tensor operations and emulate DL models with numerical processing and computing accuracy, which in effect accelerates development. TQ has already been used in collaborative research projects with the automobile industry.

1 *TensorQuant allows the automatic emulation of given TensorFlow models with any number representations of individual tensor operations.*

2 *The evalulation of the well-known ResNet-50 model shows that the concrete choice of the number representation has a considerable influence on the performance of DL applications, which is difficult to estimate in advance without the simulation in TensorQuant.*



1

ALOMA LETS GEOSCIENTISTS FOCUS ON THEIR FIELD OF EXPERTISE

1 *Presentation of a simple workflow with ALOMA: Inbound gatherers are corrected and then stacked. Input and results can be visualized interactively.*

ALOMA lifts the burden of dealing with parallelization, multi-threading, and other challenges in high-performance computing from its users. Instead, the experts for geophysical questions can focus on their area of expertise while ALOMA takes care of efficiently executing their algorithms even on large scale and heterogeneous systems.

The software is a specialized version of GPI-Space which is widely used in fields beyond geophysics such as big data and machine learning.

Complex computations on ever growing amounts of data are characteristic for the geosciences and thus geophysicists are forced to learn about HPC techniques in order to make their software run efficiently on large scale systems. We developed a system that sits in between the geophysicist and the HPC expert. Computer scientists and geophysicists together came up with ideal strategies for parallelization, data partition, and failure tolerance in the context of geophysical applications.

The heart of ALOMA, its failure tolerant runtime system to execute workflows on distributed systems, was then developed by the HPC experts of our group. For its users, the geophysicists and geologists, ALOMA is merely a black box in which they can integrate their latest developments via a well-defined interface. The learning curve for the new approach is easy to manage. Once ALOMA is installed, it takes users not more than a day to port their first module to the new system.

Quick prototype development and scaling

The main benefit of ALOMA is to quickly integrate and test newly developed algorithms and prototypes on production scale real-world problems in no time. Furthermore, existing codes and applications – even in different programming languages such as C/C++, Fortran, Matlab etc. – can be integrated as modules in ALOMA. With a graphical editor, users can combine these modules into workflows and let the software deal with the automatic parallelization and execution.

We were able to prove the feasibility of this concept in various projects with partners in the oil and gas industry, where we managed to make a customer software scale within a few days. The concept is so convincing that a Houston based company has commissioned us with switching over their existing processing software to ALOMA.

SAFECLOUDS – DISTRIBUTED INFRASTRUCTURE FOR DATA ANALYSIS IN AVIATION

The constantly growing volume of traffic poses major challenges for aviation security organisations, airports and airlines to guarantee the highest possible level of safety. Large amounts of data from various sources such as flight data recorders and radar stations are already being recorded and evaluated.

The EU project SafeClouds aims to bring together existing and new data sources in a Europe-wide infrastructure and then efficiently evaluate them using machine learning methods; the aim is to significantly improve air traffic management. A broad consortium consisting of airlines, aviation safety organisations and authorities as well as research institutions is therefore involved in the project.

Data exchange with GPI-Space

In CC High Performance Computing, we are building a multi-tier hybrid cloud infrastructure based on Amazon AWS. Our GPI-Space software is available for parallel data processing. The focus is on data and failure safety as well as easy scalability in terms of number of users, memory and computing power.

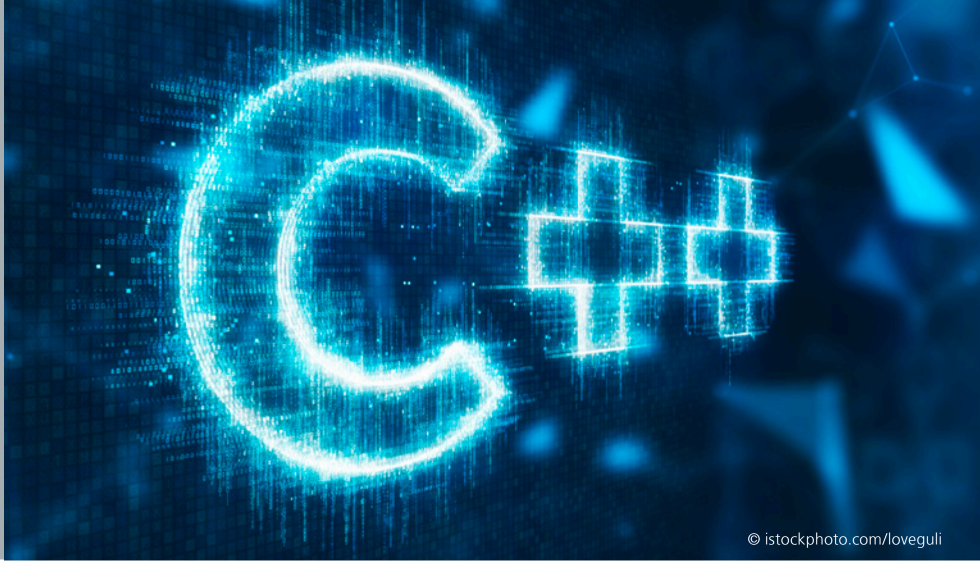
The data of the following scenarios are analyzed as an example:

- Unstable approach: A predefined range for various parameters such as altitude, speed, sink rate etc. is not maintained and can lead to a hard landing, landing abort or similar.
- Off-road safety warnings: The specified minimum amount was not reached due to geographical conditions.
- AIRPROX (Aircraft Proximity Hazard): Safety has been compromised by the minimum distance between aircraft being undercut.

The runways are also considered: the aim is, of course, their optimum capacity utilisation; in doing so, the exits to the terminals must also be taken into account while maintaining the minimum safety distances.

1 *SafeClouds analyses safety-relevant processes on the ground and in the air.*

2 *Methods of Machine Learning facilitate air traffic management.*



HIGH PERFORMANCE CONSTRAINT-PROGRAMMING

Increasingly more complex variant configurations (VC) require ever faster algorithms to be able to account for high performance, in-memory platforms like SAP S/4HANA. In a project we are offering C++ consultancy and develop cutting edge constraint solving technology and novel algorithms for SAP.

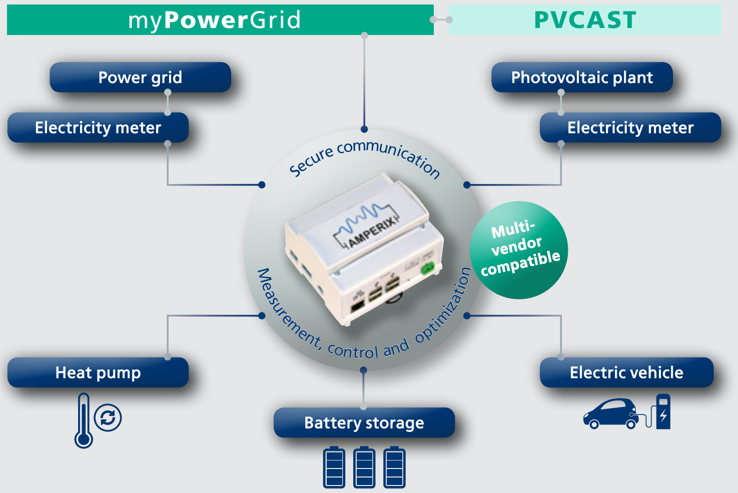
The variant configuration integrated into SAP S/4HANA offers efficient variant configuration for any kind of product lines. It does not matter if the company produces cars or pizzas or aims at merging isolated software products: All kinds of variants are described with a variant model which eventually delivers a product that fits your use case. All variants are supported cross all enterprise processes.

Constraint Solving Backend

To account for the increasing performance and scalability demands the VC Backend is ported to a leading-edge, C++-based Constraint Solving technology. This technology allows the rule-based pruning of variable domains according to the constraint model. For the definition of such models and rules we developed interfaces to be able to specify valid variable domains via variant tables and complex Boolean expressions. Further, we added new types of variables for the processing of strings and highly accurate floating-point numbers.

Challenges

The SAP Variant Configurator is used by many international customers in critical enterprise work-flows. We set the highest development standards regarding new algorithms (optimization of memory usage, compute intensity and runtime, correctness) as well as code quality (test driven development, 100% test coverage, code-reviews, continues integration on multiple platforms and compiler combinations, fuzzy testing) to be able to guarantee scalability and robustness of our solutions.



1

AMPERIX – A CENTRAL ENERGY HUB

The Amperix energy management system optimizes the operation of battery storage systems, heat pumps, and electric vehicle charging stations by providing an energy hub for home and industrial use. In conjunction with the myPowerGrid platform, it paves the way for new business models.

The Amperix records all energy flows in the household and serves as an aid for control decisions. This includes generation as well as grid demand and feed-in, while taking into account larger consumers such as heat pumps and electric vehicles. An intuitive presentation make the energy flows completely transparent for the user.

Sector coupling of electricity, heat, and electro-mobility

Initial development of the Amperix focused on battery storage systems. The evaluation of pilot installations, however, showed that a significant energy surplus occurs in the summer months despite the battery storage. Today, this surplus can, for example, heat a hot water buffer with a heat pump, or charge electric vehicles with solar power. Sector coupling of electricity, heat, and electromobility holds considerable potential for increasing the supply of locally generated energy.

Forecast-based control

Alongside the intelligent analysis of current measurement data, the Amperix also processes yield forecasts for rooftop photovoltaic systems, as well as load forecasts for household and heat pump consumption – while simultaneously factoring in the residents' wishes as to when the electric vehicle should be charged up for departure. These forecasts allow for a better use of available flexibility in the home.

Manufacturer-independent and scalable

The Amperix is vender-independent and compatible with common products. New components easily fit in the existing infrastructure. In addition, the Amperix not only reliably controls home storage devices, but also large industrial storage systems. The Amperix enables the aggregation of different systems in an internet platform, which fosters the implementation of new business models by facilitating the visualization and management of customer systems as well as the creation of virtual, large-scale battery systems.

1 *The Amperix is connected to energy meters and controllable devices on-site and to the myPowerGrid platform securely over the Internet. Intelligent control algorithms enable optimized operation of all connected devices.*

NEWS

INSECURE CHARGING INFRASTRUCTURE AT CHARGING STATIONS

Mathias Dalheimer's presentation at the annual conference of the CCC in Leipzig was very well received by the media - his topic: Safety problems at charging stations. Some loading cards are unsafe and make it possible to load on someone else's account. The weak point is the identification number; it is publicly available and can be copied at will. Communication between the charging stations and the billing back-end is also poorly protected. With little technical effort, this communication can be intercepted and the card numbers of customers can be obtained. The first shop network operators have confirmed the weak points and are taking the first steps after the great media response. A consortium of experts is also planned to systematically tackle these problems.

NEW BMBF PROJECT ON DEEP LEARNING

Hardly any other field has developed as rapidly and successfully in recent years as Machine Learning, which also includes Deep Learning. However, this procedure, which has been successful for many practical applications, requires enormous computing effort and a great deal of training data. This is why methods and infrastructures must be developed to ensure the practical predictability of increasingly complex neural networks in the future. The BMBF has been funding this since the end of 2017 in the project High Performance Deep Learning Framework – software environment for the efficient design of deep neural networks on high-performance computers with a total of 2 million euros. The project, which will run for three years, is coordinated at the Competence Center High Performance Computing.

EUROPEAN MICROPROCESSOR EPI

Together with Fraunhofer IIS, we are part of the consortium that will develop a European microprocessor by 2021. The European Processor Initiative (www.european-processor-initiative.org) focuses on high-performance computing for next generation supercomputers and embedded processing in the automotive sector. Supported by the European Union, the project is to unite competences in the field of chip development and make the EU more independent of non-European competitors in the future. The framework agreement was signed at the end of last year and development work is expected to begin in fall 2018.

Front, left to right: Farooq Arshad, Dr. Dimitar Stoyanov, Dominik Loroch, Dr. Tiberiu Rotaru, Dr. Rui Machado, Dr. Franz-Josef Pfreundt, Patrick Reh, Dr. Martin Kühn, Dr. Matthias Balzer, Dr. Abel Amirbekyan, Adrien Roussel, Frauke Santacruz, Tina Hill, Dr. Khawar Ashfaq Ahmed, Dr. Norman Ettrich, Matthias Klein, Tobias Götz, Matthias Deller, Christian Mohrbacher, Phoebe Buckheister, Bernd Lörwald, Lukas Ristau, Dr. Jens Krüger, Delger Lhamsuren, Bernd Lietzow, Thomas Olszamowski, Julius Roob, Dr. Valeria Bartsch, Dr. Mirko Rahn, Dr. Alexander Janot, Dr. Dominik Straßel, Dr. Peter Labus, Kai Krüger, Dr. Daniel Grünewald, Javad Fadaie Ghotbi, Dr. Alexander Klauer, Raju Ram, Dr. Dirk Merten