Fraunhofer SCAI

FRAUNHOFER INSTITUTE FOR ALGORITHMS AND SCIENTIFIC COMPUTING SCAI

# MACHINE LEARNING AND HPC FOR INDUSTRIAL APPLICATIONS

# HPC AND ML ENABLE NEW INDUSTRIAL APPLICATIONS

High Performance Computing (HPC) is an enabler of the recent developments and tremendous success in the field of machine learning and big data. It connects research from mathematics, computer science (including hardware and software), and application fields such as engineering, natural sciences, economics, and many more.

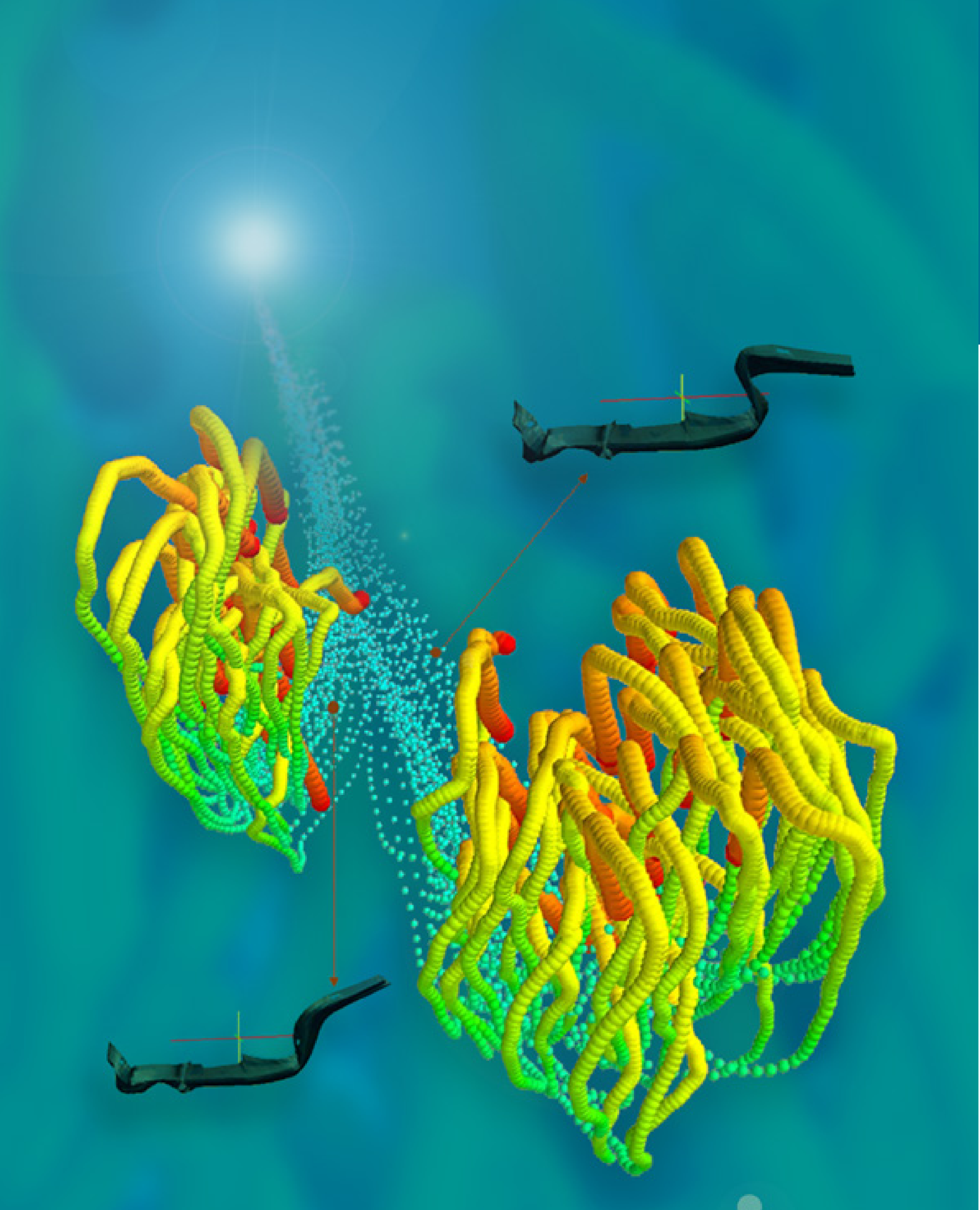
As an own research field, HPC develops scalable and reliable algorithms, efficient software systems, and new hardware solutions. It thereby covers the whole range of computer architectures, from huge clusters of interconnected computers to various kinds of specialized computing facilities. The latter includes accelerators like many-core processors, graphics processing units (GPUs), microcontrollers, or field programmable gate arrays (FPGAs) for use in machine embedded computing. The goal is always to gain optimal performance out of the given hardware, e.g. efficiency in terms of reaction time or power consumption. Furthermore, the reliability of algorithms and computed results is an essential issue.

Machine Learning (ML) stands for a revolutionary new paradigm in science and engineering. In ML, data is recognized as a driving force which leads to new knowledge and changes the way of thinking in computational sciences. For realistically-sized problems, ML and its applications require a systematic use of HPC technologies. To govern the complexity of these applications, algorithms, software and hardware have to be carefully adjusted.

Fraunhofer SCAI, based on its expertise in numerical simulation and algorithms, combines HPC and ML for breakthroughs in industrial practice. SCAI's work is focused on the development of new and intelligent algorithms and on the adaptation of data analysis methods to the specific needs of application projects. Data is not only taken from measurements or sensors, but combined with and enriched by results from physical modeling or numerical simulations and general application knowledge. Such an approach is often called a *gray box*. In particular, in engineering and natural sciences applications, this approach offers more reliability in predictive analyses and enriches the potential of ML technology. It differs from a *black box* approach which takes data *as it is* and treats it with generic data analysis methods, thereby neglecting available specific domain knowledge and structure.

SCAI offers its broad experience in computational science and engineering as well as its deep knowledge of HPC systems and programming for algorithmic developments and machine learning applications.

This brochure gives an overview on activities in HPC and ML, presenting applications as well as algorithmic approaches.



*Time-dependent low dimensional representation of hundred different numerical simulations for deformations of a car structural component.*



# NEW ALGORITHMS FOR BETTER DATA ANALYSIS

Data analysis with modern machine learning algorithms is nowadays ubiquitous in many science and engineering application domains. Based on a long history and expertise in the development of efficient numerical algorithms, Fraunhofer SCAI works on the research, development, and implementation of efficient machine learning methods with applications in several domains. Core competences of SCAI refer to, e.g.:

- efficient machine learning algorithms by dimensionality reduction or advanced numerical approaches,
- improved solutions due to enriching machine learning approaches by domain knowledge (the so-called *gray-box* approach),
- the application of machine learning to predictive simulation in science and engineering.

## Dimension reduction – a key to efficient data analysis

Instead of using standard algorithms which require huge amounts of computing power, SCAI follows a far more sophisticated approach. It is folklore in data analysis that high dimensional data typically forms lower dimensional structures. Recovering these structures is the key to efficient data exploration and analysis. SCAI develops algorithms to discover and compute such low-dimensional representations of the data. This know-how is an important advantage for customers and collaborators.

SCAI also improves the efficiency of core machine learning algorithms. A recent study has demonstrated that the integration of SCAI's software package SAMG (a highly efficient solver package for large systems of linear equations) into a machine learning package significantly reduces the run-time of the nonlinear dimensionality reduction procedures for large data sets by speeding up the required eigenvalue computation.

## The gray-box approach

Enhancing machine learning approaches with domain knowledge, such as the physical and technical processes described by the data, the gray-box approach opens up new perspectives in exploiting and gathering knowledge automatically.

Fraunhofer SCAI develops corresponding mathematical concepts which make it easily possible to integrate expected physical properties and behavior into the machine learning methods. From the mathematical point of view, a core problem here is to define a concept of distance which is suitable for the complex data arising from real-life applications.

In the end, such gray-box methods support engineers and scientists in their research and development by providing data analysis approaches which simplify their work process and allow them to concentrate on the underlying engineering and science aspects.

# COGNITIVE MACHINES WILL INTERACT WITH HUMANS

The idea behind smart factories and smart production lines is gathering data and analyzing them to be able to define actions on specific identified states. But although large amounts of data are gathered within machines, processes, and production lines in factories, very little of this data is really put to use today.

With cognitive machines and cognitive production, engineers will be able to make use of all data gathered. Machines will start to learn about their normal and abnormal behavior, allowing not only the detection of failure but also its prevention by employing predictive maintenance procedures.

Custom-fitted, sophisticated models allow machines to start to grasp their environment, learn from experience, and mimic the cognitive processes of human workers.

### Industry 4.0 means HPC

The process of learning requires data acquisition, data storage, and data analytics in the first place. Considering the substantial amount of data to be handled, there is a need for high performance computing (HPC) on small to mid-size compute clusters.

In most cases, the analytics and learning phase requires intense computation work best carried out on modern HPC systems. Fraunhofer SCAI assists partners with such needs and helps them to exploit the full potential of optimization.

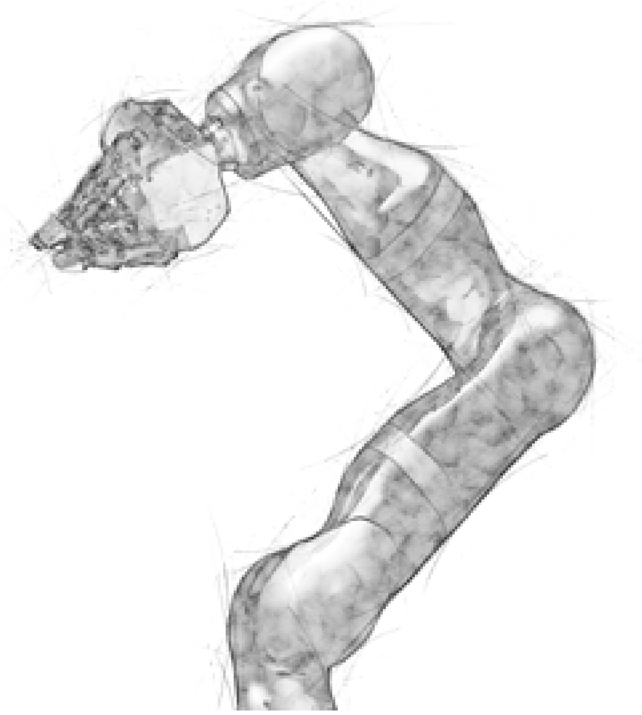
Note that the applications after the learning phase in Industry 4.0 are usually carried out by embedded system devices. The optimization challenge now is that the systems have to react in real time. SCAI provides the expertise to reduce the computational effort and to satisfy such real time constraints.

Combining HPC resources and methods with the digitalization of factories will reduce total costs of ownership, increase the productivity of the system, and improve competitiveness. SCAI provides the mathematical and computational know-how for making such ideas come true.

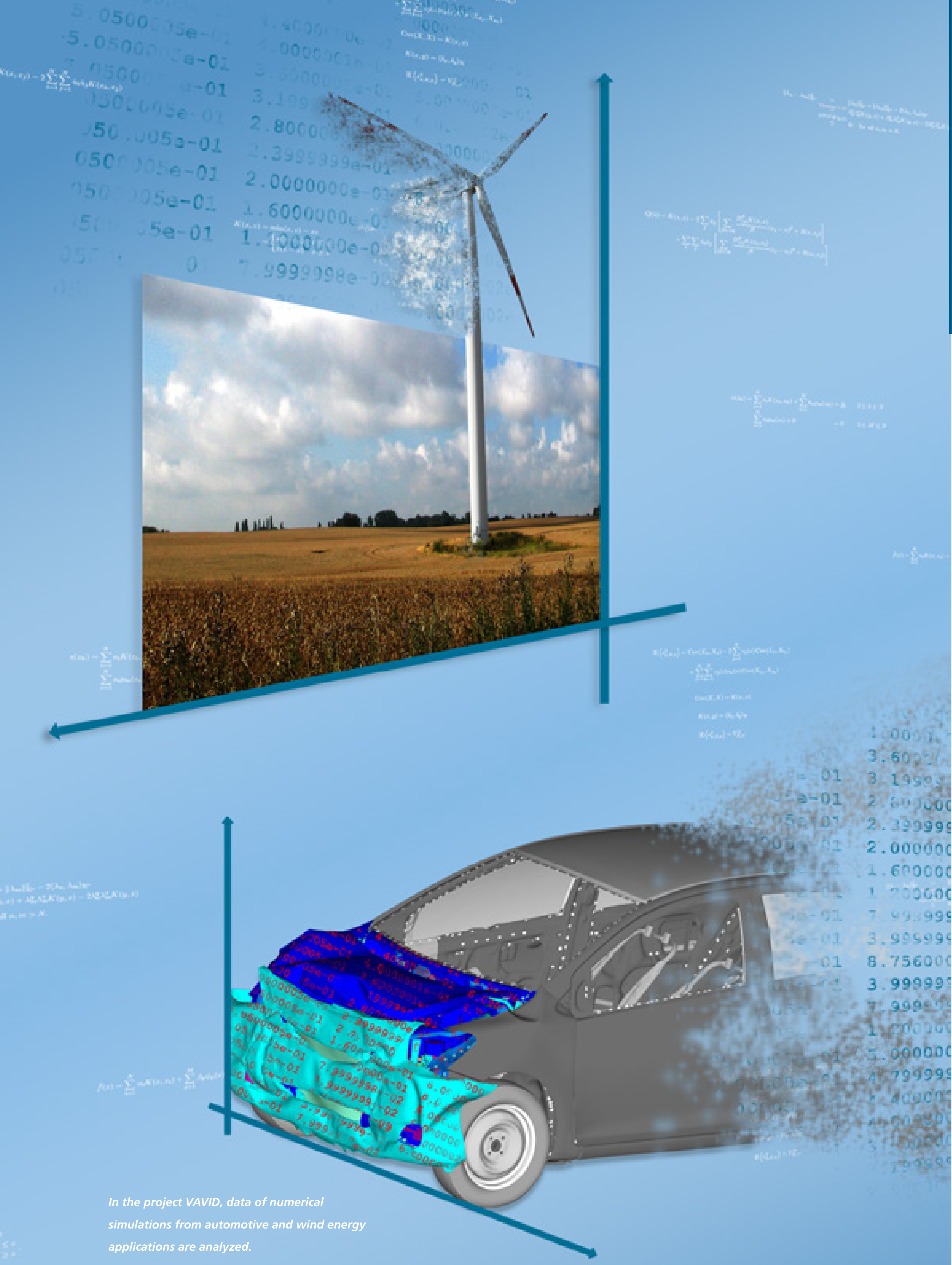
### Human-machine interaction

In the future, cognitive machines will be able to acquire a kind of self-conscience. With the support of SCAI's mathematical models and learning techniques, carefully tuned to specific use cases, those machines will be able to achieve much more, by gaining a notion of their environment.

Combining a multitude of senses, they will develop an understanding of their surroundings. This is essential for interacting with other machines and even with humans.



*In future factories humans and robots will cooperate without risk of injury for the humans.*



*In the project VAVID, data of numerical simulations from automotive and wind energy applications are analyzed.*



# ML SPEEDS UP INDUSTRIAL DEVELOPMENT PROCESSES

Today the development and improvement of new industrial products is based on complex simulation data from very detailed 3D models of a product. Such models are solved numerically using high-performance compute clusters, e.g. to simulate physical processes such as structural deformation, vibrations, or heat exchange. Product design requires the calculation of thousands of variations of a model, which might differ in their geometrical configuration, material parameters, load cases, etc. Engineering know-how is continuously demanded for investigating those changes to satisfy the functional constraints of the product as well as to fulfill safety regulations. This all congregates into a time-consuming design process which involves analysis and evaluation of the effects of those changes, which are based on engineering judgment of complex, time-dependent 3D simulations.

## Evaluation of large simulations is a bottleneck

Typically, a very detailed mesh with millions of nodes and elements evaluated at several hundred timesteps is a common representation of one simulation. Post-processing tools that visualize the model behavior or evaluate meta-data for a single simulation are used in such evaluations. The timely development of new products as well as a lack of adequate post-processing tools for the concurrent evaluation of many simulations stands in contrast to desired shorter development phases. This creates a serious bottleneck in the product development. The need to find ways of speeding up the analysis capabilities for the industry is essential.

## Reducing complexity

To address this challenge, Fraunhofer SCAI has developed new mathematical methods for the efficient analysis of thousands of complex simulations. Complexity is reduced by calculating equivalent compact representations of each simulation. To obtain such a representation, a geometric decomposition for simulations is used. In fact, this means that a few coefficients represent large variations of a detailed simulation. A time-dependent low dimensional structure is obtained, that can be used also for other tasks like optimization or for the comparison of simulations with experiments.

As a consequence, the interactive analysis and evaluation of thousands of simulations can be carried out within minutes. With the new method, a convenient parametrization of all product simulations is learned. This facilitates

- clustering simulations according to similarity,
- identification of time-dependent bifurcations,
- interpolation of new simulations, and
- the separation of physical effects.

## Applications

The described methodology has enabled the evaluation of simulation variants in the automotive industry where huge repositories of CAE (Computer Aided Engineering) models and simulation are common. Several industrial examples demonstrate its applicability for speeding up the design of new products. Application examples grow continuously as engineers realize the potential of the method.

An example is the identification of deformation modes by crash under the influence of the variation of many parameters like part thickness or material properties. With the geometric decomposition approach, complex deformations in car crash simulations are represented by few quantities that can be searched for or compared in large repositories, a task which is very difficult to accomplish using the original mesh representation. With our approach, the identification of stable deformation modes can be achieved much more easily.

In the project VAVID, funded by the German Federal Ministry of Education and Research, simulation results and the sensor data received from machines and installations are reduced to their core by using comparative analysis and data compression. Data comes from automotive and wind energy companies.

*Pipelines of a gas network – MYNTS supports the integration of classical infrastructure to a future energy backbone.*



# ML HELPS TO ACHIEVE THE ENERGY TURNAROUND

Parameter analysis and optimization are common tasks in engineering applications. SCAI develops DesParO, a machine learning software package for intuitive exploration, automatic analysis, and optimization of parametrized problems. DesParO can be coupled with simulation packages or used for measurement data. It minimizes the number of simulations needed for building up a model adaptively. Fraunhofer SCAI applies this approach in particular for energy networks.

Networks for transport of gas, electricity, water, steam or chemicals are essential for modern economy. The *Energiewende* (energy turnaround) requires optimization of energy flows in order to reduce consumption of non-renewable energy sources. The transformation of our energy backbone needs modeling, simulation and optimization of the corresponding networks.

The different energy networks can be mathematically modeled in a very similar way. Their numerical simulation can be based on the same numerical kernels. Our MultiphYsical NeTwork Simulation framework (MYNTS) is suited for all energy networks. MYNTS uses HPC technology. For parameter analysis and optimization, MYNTS provides an interface to DesParO.

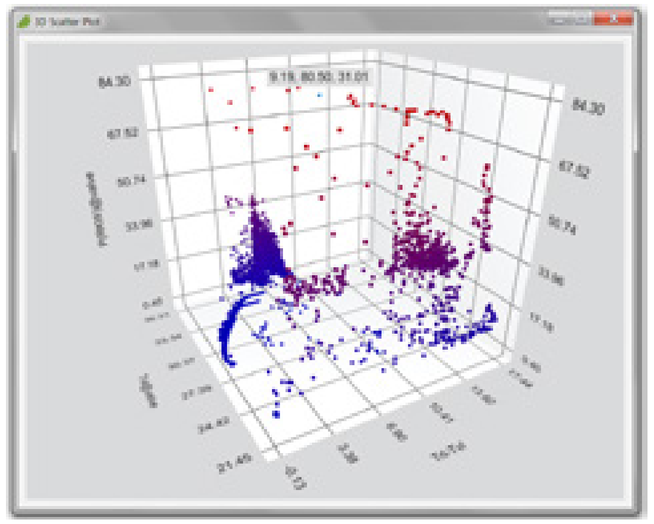
## Simulation, analysis and optimization

In order to set up an energy network model, first, the topology of the network is created. Then all the network elements are modeled. For physical devices, technical documentation is evaluated. If the behavior of a device or subsystem is not known yet, it can be learned with DesParO, based on measurement results. After network initialization, important operating scenarios have to be learned with DesParO, e.g. by clustering and/or classifying sensor data appropriately. Finally, optimization goals must be established, reflecting the intentions of relevant stakeholders.

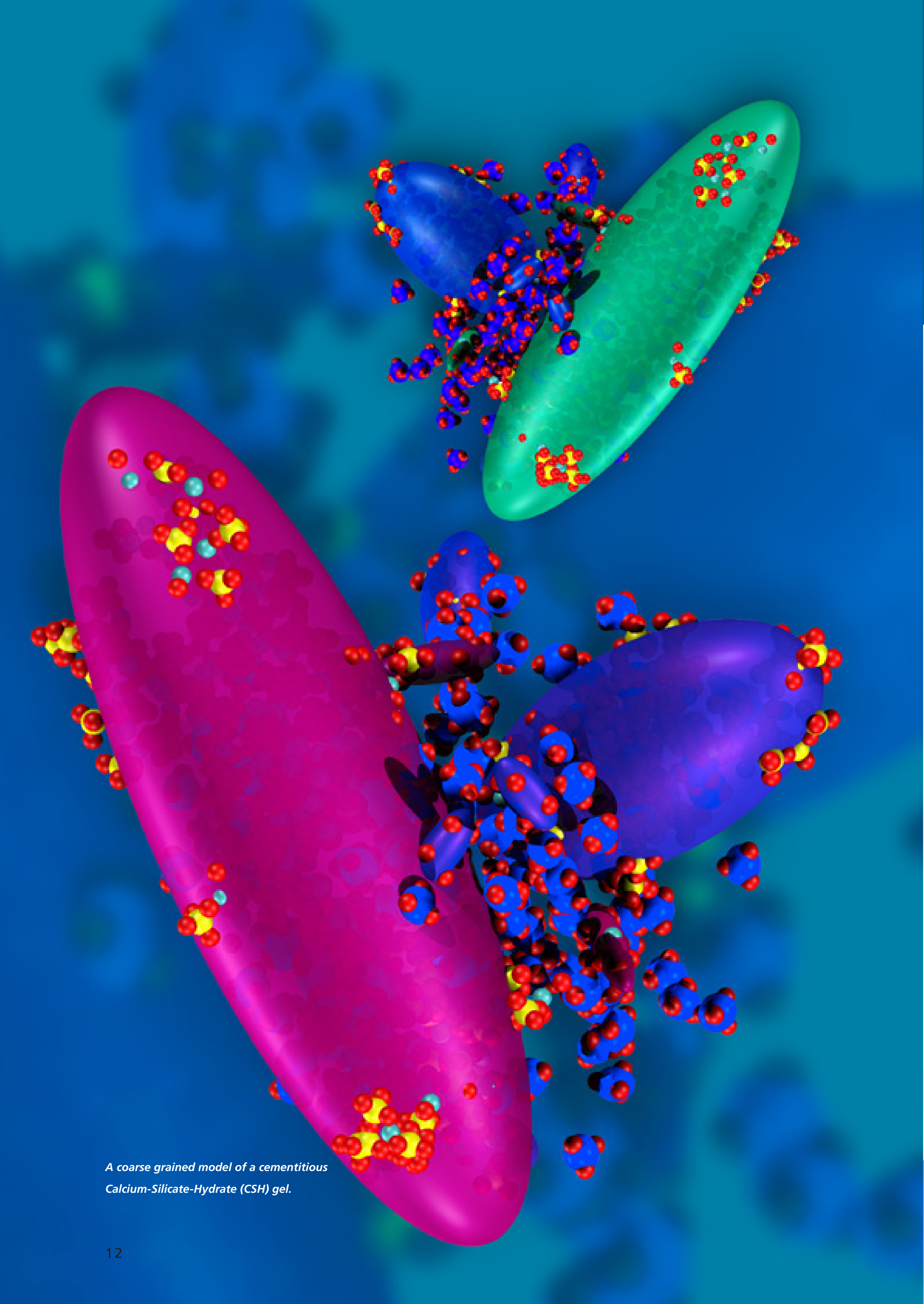
Besides simulation and optimization workflows, MYNTS offers ensemble analysis based on DesParO. So, the impact of (smaller or larger) variations of parameters can be identified. Combinations of geometrical properties, physical parameters of devices, operating scenarios or profiles of consumers, generators, or prosumers can be varied and analyzed systematically. In addition, MYNTS offers modules for physically equivalent graph reduction and for graph clustering. The latter can be used for analysis of main flow directions or comparative analysis of several operating scenarios.

## Applications

- Mathematical techniques for transforming energy networks, e.g. gas and electrical power on long-distance and distribution levels.
- Making the energy infrastructure (electrical power, gas, district heating with converters and storages) of cross-sectoral distribution grids more flexible.
- Simulation and optimization of data center energy flows.
- Energy management (DIN EN ISO 50001) and optimization for the chemical industry based on online process mining.
- Chemical micro- and macrokinetics of fuel cells.



*The three-dimensional data show the operational state of cooling an HPC center.*



*A coarse grained model of a cementitious Calcium-Silicate-Hydrate (CSH) gel.*



# ML REDUCES THE COSTS TO DEVELOP MATERIALS

Modern techniques of data analysis and machine learning open up new possibilities to use generated data for accelerated material development and process optimization – by means of experiment and simulation. Two application fields are the development of new materials with improved properties and the optimization of processes.

## New field of data-driven research and development

Besides theory, experiment and measurement – computational sciences have established themselves as a new field in research and development over the last decades. The large amount of newly-generated data from experiments and simulations together with new techniques of data analysis and the use of high-performance computers have opened the new promising field of data-driven research and development.

Fraunhofer SCAI develops and implements various methods of machine learning and data analysis in different areas of material sciences and chemistry. In particular, SCAI pursues a data-driven approach to virtual material design. For this purpose, methods of high-dimensional optimization, machine learning, data analysis, and multi-scale simulation are suitably combined. This approach is also the basis for the development of simulation-based decision support systems.

## Applications

An important application is the use of modern data analysis and machine learning to develop efficient models for predicting the properties and performance of materials. These models can be used by means of high-dimensional optimization for the efficient design of materials, molecules or processes.

In particular, SCAI develops similarity measures and kernels which are appropriate for the analysis of the chemical and materials space and uses feature extraction and dimensionality reduction techniques to extract and develop simple models for underlying complex physical processes – examples are found in the projects

- ATOMMODEL (atomic-scale modeling of novel metal-oxides in electronics)
  Development of machine learning based interatomic potentials and efficient models for predicting materials and molecular properties
- MultiModel (multi-scale simulations software for modeling of growth and diffusion in materials systems)
  Development and application of efficient similarity measures and kernels for graphs and atomic environments
- SchmiRMal (molecular modeling and simulation of ionic liquids, octanol/water and membrane/water partition coefficients)
  Estimation of the toxicity of chemicals by means of machine learning and data analysis
- Human Brain Project (integration methods for multi-simulator multi-scale simulations)
  Machine learning based surrogate models for spiking neural network models

Altogether, innovative data-driven virtual design approaches open the possibility to substantially reduce costs to develop new technologies.

# WORKSHOPS

### Workshops on machine learning and data analysis

SCAI organizes training sessions and workshops on machine learning techniques that have an introductory character and/or address advanced methods. Topics include the data processing cycle of data understanding and preparation, data analysis with methods from machine learning, up to the use and deployment of the analysis results.

In addition to general training, the contents of the workshops can also be adapted to the requirements and interests of the customers.

www.scai.fraunhofer.de/workshop-maschinelles-lernen

### Training programs of the Fraunhofer Big Data Alliance

SCAI is also involved in the professional education of data scientists and big data specialists of the Fraunhofer Big Data Alliance:

www.bigdata.fraunhofer.de/de/datascientist.html



# CONTACTS

Fraunhofer Institute for Algorithms and
Scientific Computing SCAI
Marketing and Communications
Schloss Birlinghoven 1
53757 Sankt Augustin, Germany
Phone +49 2241 14-4444
muk@scai.fraunhofer.de
www.scai.fraunhofer.de

### Image Sources

Cover: Fotolia.com – Nataliya Hora
Page 6: Fotolia.com – zapp2photo
Page 10: MEV-Verlag

All other pictures: Fraunhofer SCAI



Machine Learning Algorithms and Applications
Prof. Dr. Jochen Garcke
Phone +49 2241 14-4015
jochen.garcke@scai.fraunhofer.de

HPC and Cognitive Machines
Dr. Thomas Soddemann
Phone +49 2241 14-4076
thomas.soddemann@scai.fraunhofer.de

Energy Networks
Prof. Dr. Tanja Clees
Phone +49 2241 14-4074
tanja.clees@scai.fraunhofer.de

Material Design
Dr. Jan Hamaekers
Phone +49 2241 14-4230
jan.hamaekers@scai.fraunhofer.de