Fraunhofer Institute Centre Schloss Birlinghoven

- largest research centre for informatics and applied mathematics in Germany
- around 700 employees, thereof 500 scientists, approx. 200 students and trainees
- academic links
  - U Bonn
  - U Cologne
  - RWTH Aachen
Fraunhofer SCAI

- numerical simulations in the r&d of products and processes
- fast solver technologies, high performance computing
- optimization in packaging, logistics, production and design
- data analysis for
  - chemical and biological databases (semantic text analysis)
  - engineering data (numerical simulations, sensor measurements)
  - retail (online shops, insurance companies)
- energy network simulation, analysis and optimization
- virtual material design (material science, nanotechnology)
- computational finance
- employees (2014): ca. 100 (plus ca. 40 students)
- budget (2014): ca. 10 million Euros, from industry: ca. 50%
Fraunhofer Big Data Allianz

- founded 2014 by 25 Fraunhofer institutes
- organised along business / research areas
  - production & industry 4.0
  - logistics & mobility
  - life sciences & health care
  - energy & environment
  - security
  - business & finance
- joint offering of training for big data scientists
- workshops with industry to identify needs and potential
- many companies think about their big data strategy
- see danger of worthwhile data centralised at US companies
Big Data at Fraunhofer SCAI

- Bioinformatics
  - Information and knowledge extraction in the life sciences
  - Applied chemoinformatics: prediction of new drug candidates
  - Grid and cloud infrastructures for the life sciences ("cloud4health")

- High-Performance Analytics
  - Analysis, simulation, and optimization for energy networks

- Numerical Data-Driven Prediction
  - Analysis of time series data from condition monitoring systems
  - Algorithms for data analysis in online retail

- High-Performance Computing
What is Big Data for Numerical Simulation Data?

- Numerical simulation used in many industries and sciences
  - Numerical weather forecasts
  - Automotive engineering – crash simulation applications
  - Wind turbine design - behaviour under lots of different winds
  - Wind farm design – computational fluid dynamics simulation
  - Oil- and gas reservoir simulation
  - ...

- Typical goal: Analyse influence of parameters on behaviour
- Common analysis: Look only at key output values
- Big data approach: Each full simulation run is data instance
  - Generates huge amounts of data / repository of simulations
    - Have to be processed, handled and stored appropriately
    - Need to be investigated and analyzed in an interactive fashion
Numerical Simulation of Car Crashes

- virtual product development Computer-Aided Engineering (CAE) follows Computer-Aided Design (CAD)
- car modell is covered by mesh (discretized by a grid)

- for each grid point we know and use equations describing
  - equation of motion for:
    - acceleration / velocity / displacement
  - outer forces (obstacle)
  - inner forces (deformation of material)
Data from Numerical Simulations of a Car Crash

- ca. one million points (physical distance in millimeters)
- time steps of numerical simulations in micro-seconds
- realistic simulation take hours on HPC-clusters
- data volume:
  number of simulations × grid points × saved time steps
  - a full simulation is in $\mathbb{R}^{100,000,000 \times 3}$
  - a time step of a simulation is in $\mathbb{R}^{1,000,000 \times 3}$
  - a physical part in a time step of a simulation is in $\mathbb{R}^{50,000 \times 3}$
Several Results of Crash Simulations
Several Results of Crash Simulations
Several Results of Crash Simulations

Slot 2: TOYOTA YARIS MODEL (NCAC V01) - State 14 at time 0.065000
Several Results of Crash Simulations
Several Results of Crash Simulations
Overlay of Two Crash Simulations
Mathematical Data Analysis for Crash Simulations

- typical design quantities studied by engineers in scenarios
  - plate thickness
  - geometric changes
  - material properties
  - material modelling

- typical key output quantities are of user interest
  - firewall intrusion or other displacements
  - Head Injury Criterion (HIC) index or other computed criteria
  - time curves

- typical goal: determining trends or correlations
- currently constrained to input numbers and output numbers
- no tools for geometric input variations or deformations
- our aim automatic organisation of raw numerical simulation data from several scenarios
Handling of Bundles of Numerical Simulation Data

- for analysis of simulations, data needs to be easily accessible
- concerns efficient transfer, storage, and access
- simulation data is bulky data, therefore
  - not stored in database but in special file formats
  - need to organise simulation results “database-like” in SDM
  - store meta data, derived data, etc. with simulation data (HDF5)
- employ data storage server
  - compress simulation data (storage, transfer, visualisation)
  - compute mainly at data, not at client
  - exploit HPC computational capabilities of server
- if possible integrate sensor and measurement data
  - pre-processing for aligning real and simulation data
Big Data Infrastructure and Processing Modules

- big data
  - models
  - simulations
  - measurements

- HPC database system
  - preprocessing
  - compression
  - data extraction

- extracted data
  - derived data
  - meta data

- data manager
  - computationally expensive
  - data analysis tools
  - simple

- visualisation
  - graphic user interface
  - visualisation

- interactive big data analysis

supported by BMBF Big Data initiative

jochen.garcke@scai.fraunhofer.de
Copyright 2015 Fraunhofer Gesellschaft
Grouping of Numerical Simulation Data

- we study machine learning techniques to allow
  - simpler engineering by interactive navigation of variants
  - detecting patterns in bulky simulation data
  - detailed study of bulky simulation data for further insights
- find nearby simulation data
  - similar geometric deformations
  - similar production curves (oil, energy from wind)
(Nonlinear) Dimensionality Reduction

- need to find **data-driven distance** concept
  - suitable for high-dimensional data
  - suitable for application
- represent $d$-dimensional data in $s$-dimensional space, $d \gg s$
- goal: find **intrinsic dimension** $s$ of simulation vectors
- **nonlinear** dimensionality reduction / manifold learning
Study on Position of Bumper for Toyota Yaris
Main Structure of Interest

- deformations of grid points of this structure are our data
- embed into 3 dimensions by preserving data-driven distance
- 250 simulation runs, using ca. 100,000 points per run
Three Stable Parameter Regimes
Three Stable Parameter Regimes
Three Stable Parameter Regimes
Typical Behaviour per Mode
Typical Behaviour per Mode
Typical Behaviour per Mode

r=20 mm
Conclusions

- big data for numerical simulation data implies
  - efficient storage, and access of raw and meta data
  - current and new tools for efficient post-processing analysis
- make life for engineer easier
  - less time for detailed data analytics
  - investigate more information from numerical simulation results
- research: “good” distance measures suitable for specific
  - application domain and
  - analysis goal
- long term aim: detailed concurrent analysis of numerical simulation data and measurement data

jochen.garcke@scai.fraunhofer.de Copyright 2015 Fraunhofer Gesellschaft