# Comparative Analysis of Engineering Measurements and Simulation Data

Jochen Garcke

Fraunhofer SCAI

universität bonn

March 2016

# Fraunhofer Institute Centre Schloss Birlinghoven

- research centre for applied mathematics and informatics
- around 700 employees, thereof 500 scientists, approx. 200 students and trainees
- academic links
  - U Bonn
  - U Cologne
  - RWTH Aachen
  - Hochschule Bonn-Rhein-Sieg

# 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)
  - virtual product development (numerical simulations)
  - condition monitoring (sensor measurements)
  - robust design
- energy network simulation & optimization (inc. HPC-setups)
- virtual material design (material science, nanotechnology)
- computational finance
- employees (2015): ca. 100 (plus ca. 30 students)
- budget (2015): ca. 10 million Euros, from industry: ca. 46%

# VAVID - Vergleichende Analyse von ingenieurrelevanten Mess- und Simulationsdaten

Consortium

- Bosch Rexroth Monitoring Systems
- GE Global Research
- GNS mbH
- SCALE
- SIDACT
- TU Dresden (ZIH & Database Systems Group)

Associated partners

- Audi, ParStream/Cisco, Volkswagen

VAVID supported by the BMBF Big Data Initiative

# Big Data for Numerical Simulations

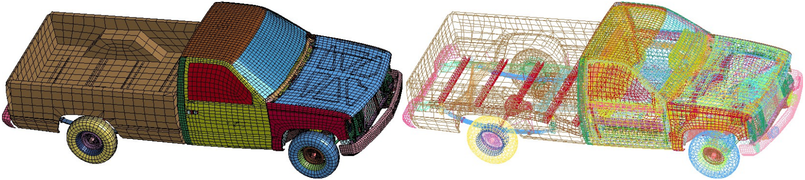
- numerical simulation used in many industries and sciences
  - automotive engineering – crash simulation applications
  - wind turbine design - behaviour under lots of different winds
  - wind farm design – computational fluid dynamics simulation
  - oil- and gas reservoir simulation
  - ...
- typical goal: analyse influence of parameters on behaviour
- common analysis: look only at key output parameters
- big data approach: each instance is a full simulation run
- can be very high dimensional data, e.g. $\sim 10^8$ (time $\times$ nodes)
- often hundred simulation runs (or more)
- 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
- reality check: combine simulation data with “real” data

# Crashtest at Fraunhofer EMI Crash Facility

# Crashtest at Fraunhofer EMI Crash Facility

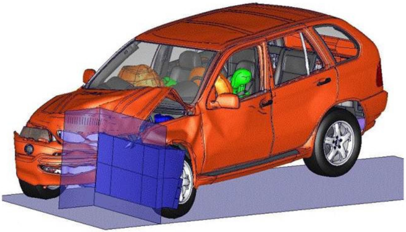
# Numerical Simulation of Car Crashes

- in 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 node we know and use equations describing
  - equation of motion for:
    acceleration / velocity / displacement
  - outer forces (obstacle)
  - inner forces (deformation of material)
- several commercial solvers in use at car companies

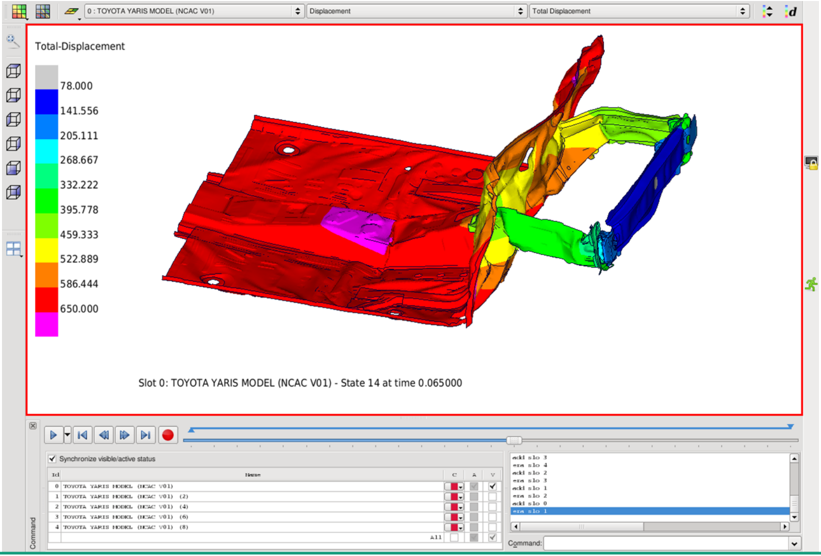
# Data from Numerical Car Crash Simulations



- ca. one million nodes (physical distance in milimeters)
- time steps in micro-seconds, about one hundred saved
- realistic simulation take hours on HPC-clusters
- data volume:
  number of simulations $\times$ grid points $\times$ 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 $\mathbb{R}^{50.000 \times 3}$

# Numerical Simulation

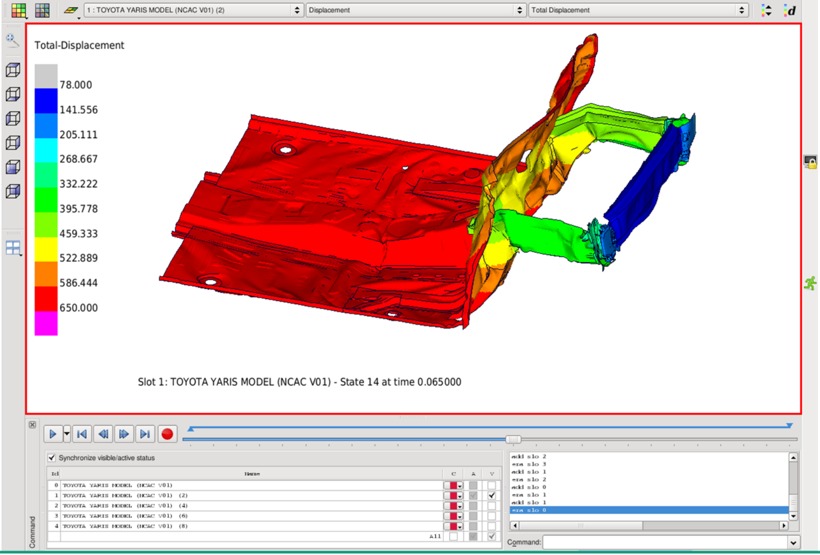
# Several Results of Crash Simulations



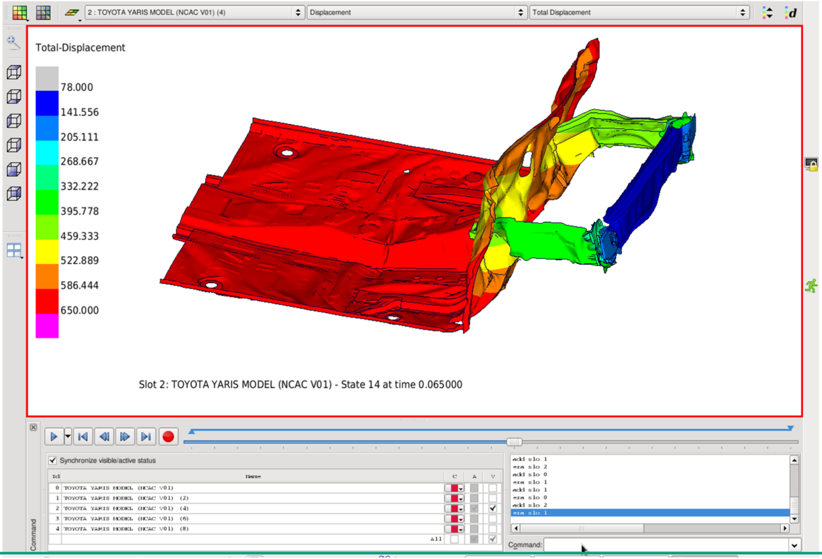

jochen.garcke@scai.fraunhofer.de

Fraunhofer SCAI

# Several Results of Crash Simulations



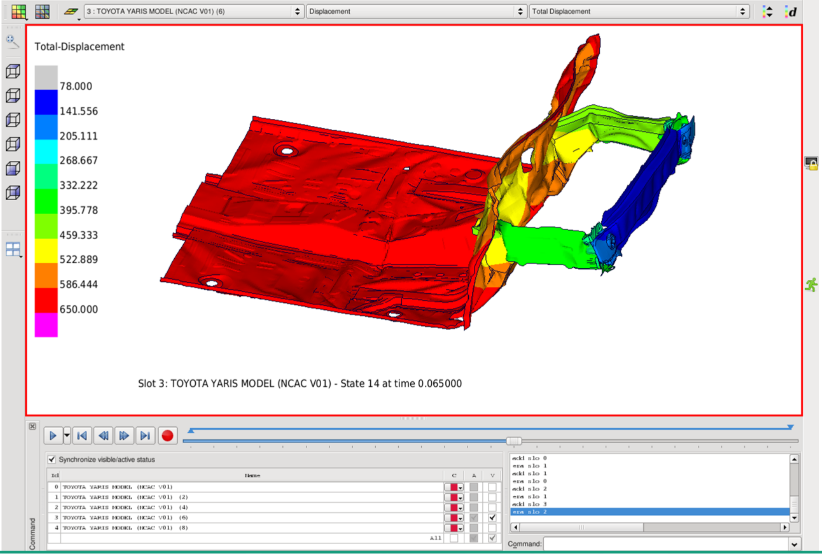
# Several Results of Crash Simulations





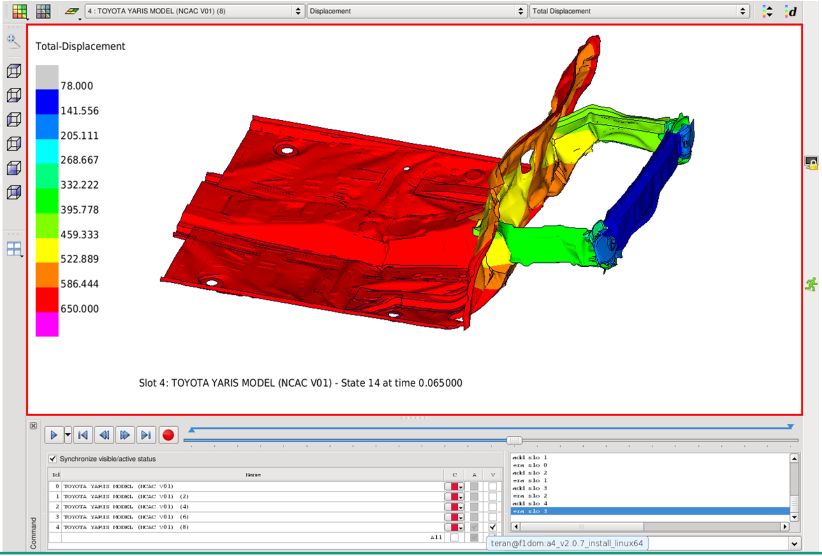

Fraunhofer SCAI

# Several Results of Crash Simulations

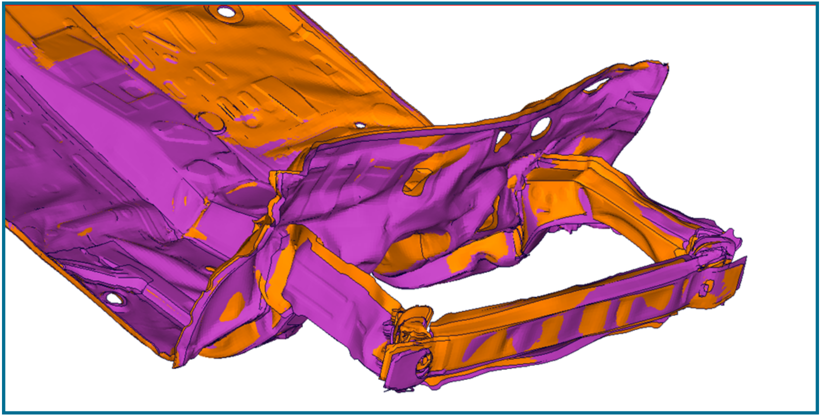



jochen.garcke@scai.fraunhofer.de

# Several Results of Crash Simulations



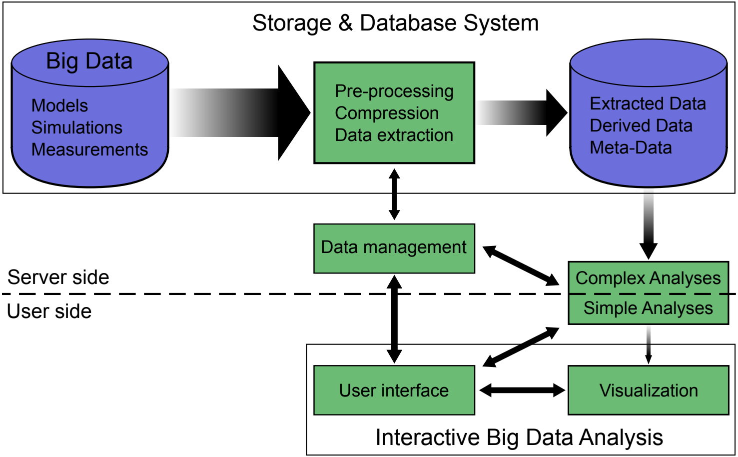
# Overlay of Two Crash Simulations



# Handling of Bundles of Numerical Simulation Data

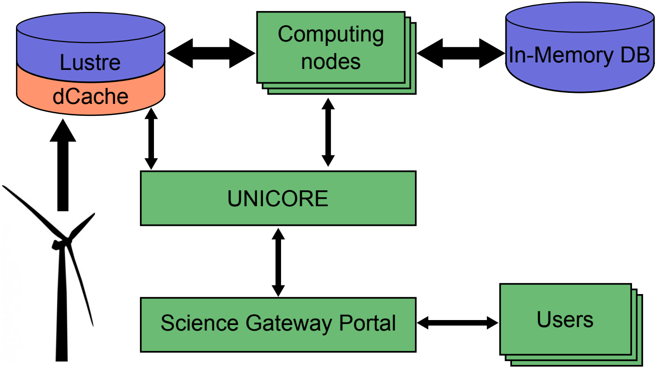
- for analysis of simulations, data needs to be easily accessible
- concerns efficient transfer, storage, and access
- simulation data is bulky data, therefore
  - not stored in database but in special file formats
  - need to organise simulation results “database-like” in SDM
  - store meta data, derived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

# Big Data Processing Modules




VAVID supported by the BMBF Big Data Initiative

# Proposed Architecture for Rotor Blade Sensor Data




- aimed for sensor information from up to 600 wind turbines
- raw data stream about 100 MiB per hour per wind turbine
- data highly fragmented, millions of single measurements
- sensor data is to be kept for about 20 years

# Data Analysis for Crash Simulations

- typical design quantities studied by engineers in scenarios
  - plate thickness
  - geometric changes
  - material properties
  - material modelling, ...
- typically 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 or output numbers
- no tools for geometric input variations or deformations
- our aim: automatic organisation of several numerical simulation results

# Grouping of Data

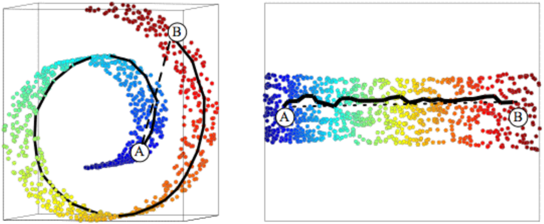
- we study machine learning techniques to allow
  - simpler engineering by interactive navigation of variants
  - detecting patterns in bulky simulation data
  - detailed study of bulky simulation data for further insights
- find nearby simulation data
  - similar geometric deformations
  - similar time curves (behaviour of simulated wind turbines)

# Simulation Data is Not Really High-Dimensional

- let us very strongly simplify things
- consider 1D-grid $[r_1, r_2, \dots, r_d]$, describing deformation
- only material parameter $\alpha$ is changed, compute $N$ variations
- simulation result $S_{\alpha_i}$ as $x_i = S_{\alpha_i}(r_1, \dots, r_d) \in \mathbb{R}^d, i = 1, \dots, N$
- simulations $x_i$ are on an intrinsic one-dimensional curve !
- $u(\alpha) : \mathbb{R} \to \mathbb{R}^d, x_i = u(\alpha_i)$ transfers material parameters $\alpha_i$ into high-dimensional simulation space
- but we do not know $u(\alpha)$, we only have data $S_{\alpha_i}$
- we now go backwards:
  try to approximately reconstruct $u(\alpha)$ from data similarity

# (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




jochen.garcke@scai.fraunhofer.de

# Arrange Data on Graph for Manifold Learning

- in last 15 years several spectral approaches use steps
  1. build graph from local neighbourhood in input space
  2. construct matrix based on chosen mapping and geometry
  3. spectral embedding via top or bottom eigenvectors of matrix
- (all) spectral approaches to manifold learning crucially rely
  - on neighbourhood graph of data ($k$-nn or $\epsilon$-neighbourhood)
  - Euclidean embedding of this graph
- one can define
  - graph metric: $\text{dist}_G(i,j)$ length of shortest path from $i$ to $j$
  - distanz matrix: $D^G$ with entries

$$d_{ij}^G := \text{dist}_G^2(i,j), \ i,j \in V$$

# IsoMap [Tenenbaum et al., 2000]

- use classical MDS-approach with new metric
- replace Euclidean distance by geodesic distance on manifold
- approximate geodesic distance by shortest path in graph

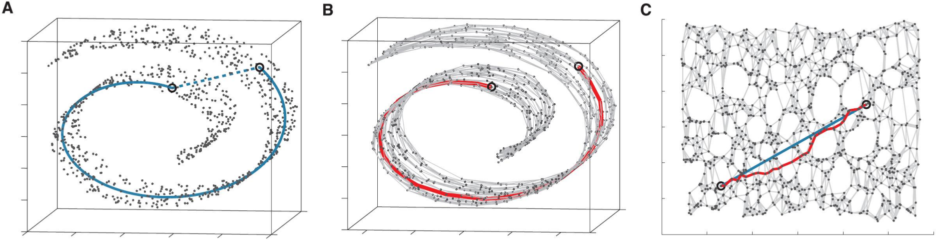
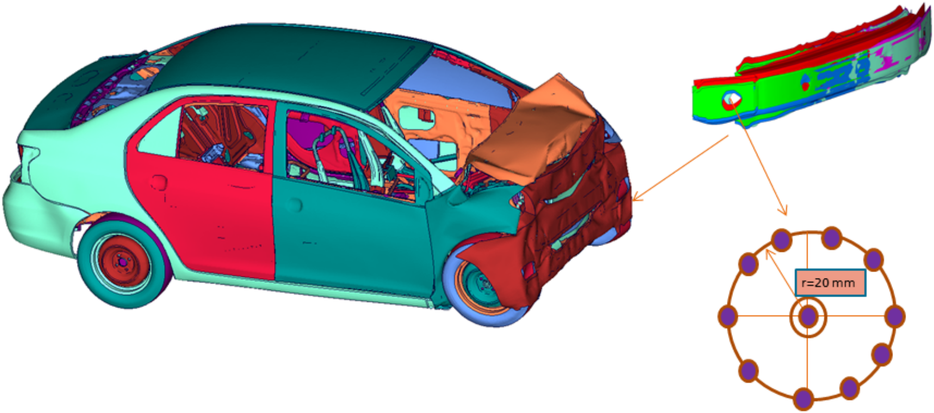


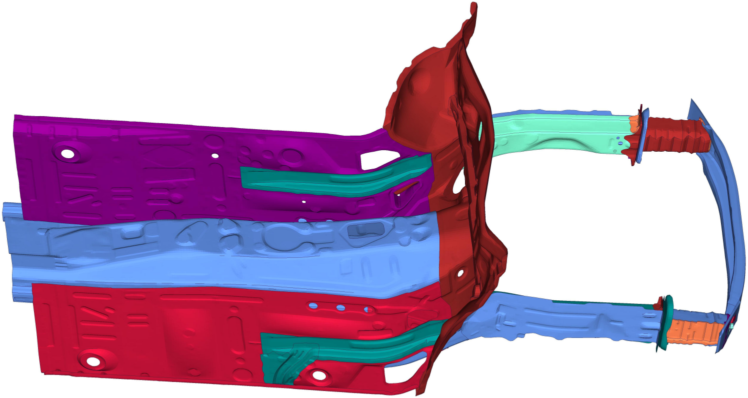

**Fig. 3.** The "Swiss roll" data set, illustrating how Isomap exploits geodesic paths for nonlinear dimensionality reduction. (**A**) For two arbitrary points (circled) on a nonlinear manifold, their Euclidean distance in the high-dimensional input space (length of dashed line) may not accurately reflect their intrinsic similarity, as measured by geodesic distance along the low-dimensional manifold (length of solid curve). (**B**) The neighborhood graph *G* constructed in step one of Isomap (with $K = 7$ and $N = 1000$ data points) allows an approximation (red segments) to the true geodesic path to be computed efficiently in step two, as the shortest path in *G*. (**C**) The two-dimensional embedding recovered by Isomap in step three, which best preserves the shortest path distances in the neighborhood graph (overlaid). Straight lines in the embedding (blue) now represent simpler and cleaner approximations to the true geodesic paths than do the corresponding graph paths (red).

- image taken from [Tenenbaum et al., 2000]

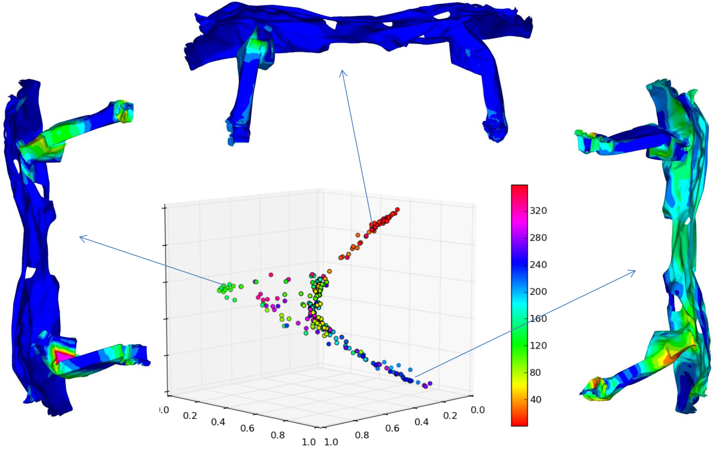
# Study on Position of Bumper for Toyota Yaris



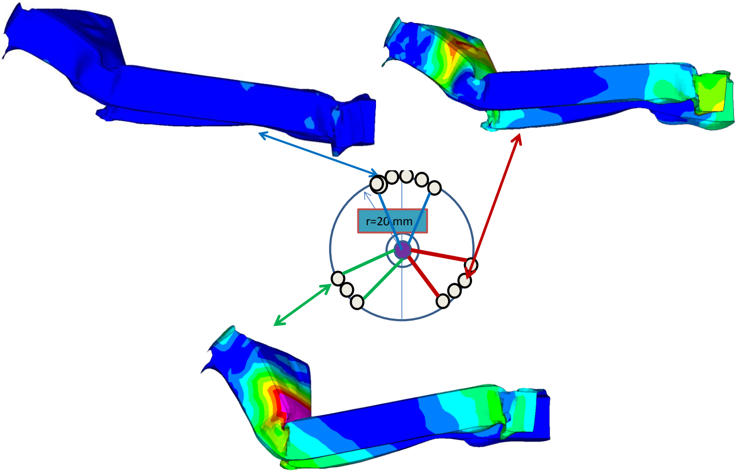
# Main Structure of Interest



# Embedded Simulations - 3 Dominant Modes



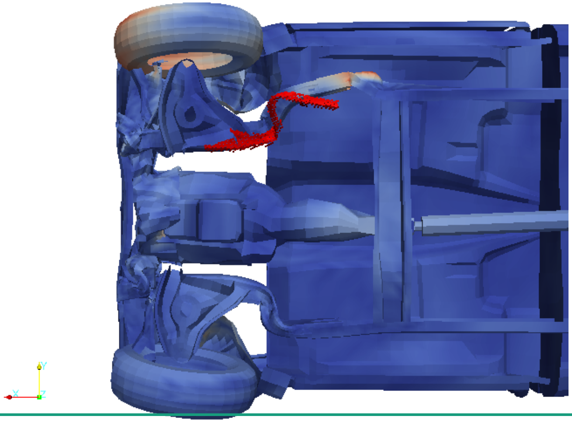
# 3 Dominant Modes - Corresponding Angle Range




jochen.garcke@scai.fraunhofer.de

# Ongoing: Match Between Reality and Simulation

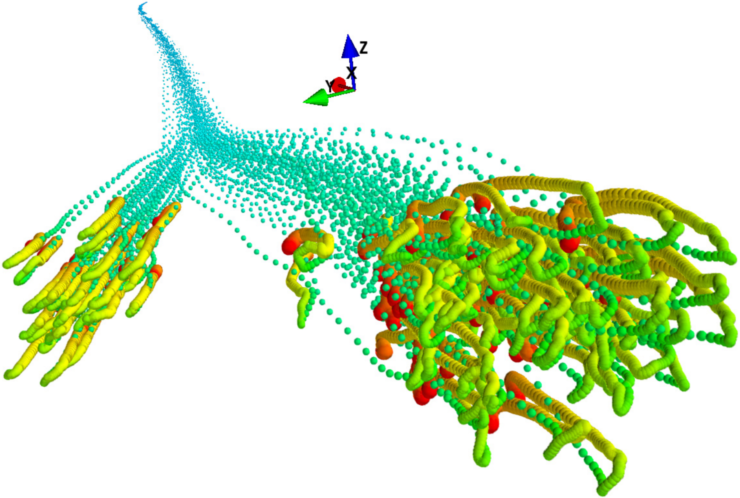
- analyse 3*D*-video of car crash (with Fraunhofer EMI and IOF)
- matching (or best) simulation for measured 3D-point cloud



# Operator Dependent Spectral Basis

- spectral decomposition of surfaces / manifolds
- work in thereby computed spectral basis
- based on operators invariant to a (chosen) transformation
- work in quotient space of deformations / transformation group
- reduced multiscale representation, depends on operator (and data)
- obtained spectral coefficients give data driven distance measure [Iza-Teran and G., 2015]

# Bifurcation in Crash Behavior, Visualised in Time



# Mode 1 - Translation

# Mode 2 - Rotation

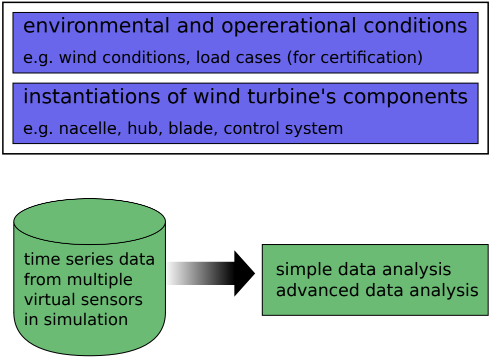
# Mode 3 - Global Deformation

# Mode 4 - Local Deformation

# Combined Modes

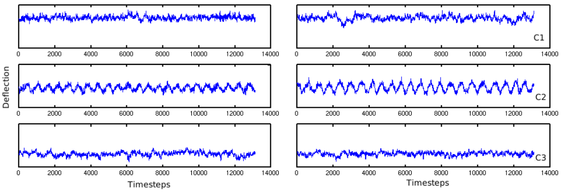
# Morph Simulation Beam to Observed Point Cloud

# Time Series Data from Wind Turbine Simulations

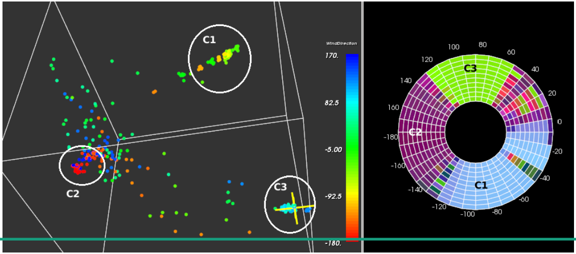



- numerical simulations of wind energy plants during
  - design
  - modification / upgrade
  - certification
  - site-dependent load assessment

# Data Analysis for Wind Turbine Simulation Data




- use nonlinear dimensionality reduction for simplified
  - detection and analysis of anomalies
  - identification and analysis of similar behaviour




jochen.garcke@scai.fraunhofer.de

# Summary

- big data for numerical simulation data implies
  - efficient storage and access for raw and meta data
  - current and new tools for efficient post-processing analysis
- math view: find "good" distance measures for specific
  - application domain and
  - analysis goal
- matching of simulation and measurement data
- techniques transferable to other numerical simulation data

# References

Rodrigo Iza-Teran and Jochen Garcke. Operator based multiscale analysis of simulation bundles. 2015. submitted.

J. B. Tenenbaum, V. de Silva, and J.C Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(5500):2319–2323, 22 December 2000.